# Inference on Factor Structures in Heterogeneous Panels

Carolina Castagnetti
University of Pavia

Eduardo Rossi
University of Pavia

Lorenzo Trapani
Cass Business School, City University London

November 5, 2013

**Abstract**

This paper develops an estimation and testing framework for a stationary large panel model with observable regressors and unobservable common factors. We allow for slope heterogeneity and for correlation between the common factors and the regressors. We propose a two stage estimation procedure for the unobservable common factors and their loadings, based on Common Correlated Effects estimator and the Principal Component estimator. We also develop two tests for the null of no factor structure: one for the null that loadings are cross sectionally homogeneous, and one for the null that common factors are homogeneous over time. Our tests are based on using extremes of the estimated loadings and common factors. The test statistics have an asymptotic Gumbel distribution under the null, and have power versus alternatives where only one loading or common factor differs from the others. Monte Carlo evidence shows that the tests have the correct size and good power.

**JEL codes**: C12, C33.

**Keywords**: Large Panels, CCE Estimator, Principal Component Estimator, Testing for Factor Structure, Extreme Value Distribution.

# 1 Introduction

Consider the following model for stationary panel data:

$$y_{it} = \beta_i' x_{it} + \gamma_i' f_t + \epsilon_{it}, \tag{1}$$

$$x_{it} = \Lambda_i f_t + \epsilon_{it}^x, \tag{2}$$

where $i = 1, ..., n$, $t = 1, ..., T$, $x_{it}$ is an $m$-dimensional vector of observable explanatory variables and $f_t$ is an $r$-dimensional vector of unobservable common factors; in equation (2), $\Lambda_i$ is a matrix of coefficients of dimension $m \times r$. Model (1)-(2) is based on Pesaran (2006), and it arguably has a huge potential for empirical applications. In the context of finance, $y_{it}$ could represent the excess return on an asset; then, as pointed out by Bai (2009a), $f_t$ could represent a vector of unobservable factor returns, which are added to the observable ones (e.g. the Book-to-Market ratio) that are typically employed. Kapetanios and Pesaran (2007) consider an APT model allowing for individual asset returns to be affected by common factors (both observable and unobservable). In a similar setup, Castagnetti and Rossi (2013) adopt a heterogeneous panel with a multifactor error model to study the determinants of credit spread changes in the Euro corporate bond market. Factor models are also useful in the context of estimating production functions, where $x_{it}$ is a set of observable factor inputs, and $f_t$ allows to consider cross sectional dependence as arising from common shocks or e.g. spillover effects determined by policy or technology shocks. For example, Eberhardt and Teal (2012) adopt a common factor model approach to estimate cross-country production functions for the agriculture sector. Similarly, Eberhardt, Helmers and Strauss (2013) consider the impact of spillovers in the estimation of private returns to R&D allowing for a common factor framework. Another promising field of application is the prediction of mortality rates (or their first difference), where the seminal Lee-Carter model (Lee and Carter, 1992) has been extended to incorporate idiosyncratic explanatory variables as well as the traditional factor structure - see French and O'Hare (2013) and the references therein.

As far as conducting inference on (1) is concerned, the inferential theory on the slope coefficients $\beta_i$ has been developed in various contributions. Particularly, Pesaran (2006) proposes a family of estimators for $\beta_i$ based on instrumenting the $f_t$s through cross sectional averages of the $x_{it}$ and $y_{it}$; such estimation techniques are referred to as the Common Correlated Effects (CCE) estimators. One of the key features of the CCE estimator is that it does not require any inference to be carried out on $\gamma_i$ or $f_t$. Pesaran and Tosetti (2011) and Castagnetti and Rossi (2013) show that, in principle, residuals computed from (1) using CCE estimators can be used to extract $\gamma_i$ and $f_t$ using e.g. Principal Components (henceforth, PC). However, the properties of the estimated $\gamma_i$ and $f_t$ are not discussed. In addition to the CCE estimators, Bai (2009a) develops a different estimation technique for (1)-(2) under the assumption of homogeneous slopes, i.e. $\beta_i = \beta$. Such technique is known as the Interactive Effect (henceforth IE) estimator, and it is based on iteratively computing $\beta$ for given values of $\gamma_i$ and $f_t$, and then $\gamma_i$ and $f_t$ for a given value of $\beta$. Although results are available for the estimated triple $(\beta, \lambda_i, f_t)$, inference is developed under the assumption of homogeneous $\beta_i$s; moreover, no explicit asymptotics for $\gamma_i$ or $f_t$ is derived beyond consistency. Despite this, inference on $\gamma_i$ and $f_t$ is likely to be important in many settings. For instance, where a multifactor error structure is employed for the purpose of dimension reduction, or simply when explanatory variables may not be observable. In such cases, it could be relevant to know whether there is indeed a factor structure in (1), or whether common effects can be adequately represented by more parsimonious models such as a model with cross-sectional or time dummies, as also studied by Sarafidis, Yamagata and Robertson (2009), and Bai (2009a) in the context of model (1) with homogeneous slopes. In this case, the asymptotics of the estimated common factors and loadings is obviously a first, fundamental step in order to construct tests for the presence of a multifactor error structure.

This paper makes two contributions to the literature. Firstly, we derive the inferential theory for the unobservable common factors $f_t$ and their coefficients $\gamma_i$ in (1)-(2). We estimate $\gamma_i$ and $f_t$ by applying PC to the residuals computed from (1) using the CCE estimator. This two-stage procedure builds on an idea of Pesaran (2006, p.1000), and Pesaran and Tosetti (2011), while the asymptotics of the estimated $(\gamma_i, f_t)$ is studied by adapting the method of proof in Bai (2009a) to the case of heterogeneous $\beta_i$s.

Secondly, we develop two tests: one for the null that $\gamma_i = \gamma$ for all $i$, and one for the null that $f_t = f$ for all $t$. The rationale for these two tests can be understood by noting that, as Pesaran (2006) points out, model (1)-(2) nests various alternative specifications. In the case of homogeneous loadings (i.e. $\gamma_i = \gamma$), equation (1) is tantamount to a panel regression with a time effect - therefore there is no real common factor structure. This fact is used by Sarafidis, Yamagata and Robertson (2009) to test for cross dependence in a dynamic panel context. Similarly, in the case of homogeneous factors (i.e. $f_t = f$), equation (1) boils down to a heterogeneous panel with individual effects - in this case, too, there is no real common factor structure. Therefore, the two tests described above can be used to verify whether a factor structure in (1)-(2) indeed exists, or whether simpler specifications nested in (1)-(2) should be employed. Both tests should therefore be employed *before* trying to estimate any factor structure, including the number of common factors, as we also discuss in Section 3. In this respect, our paper is related to a recent contribution by Baltagi, Kao, and Na (2012), who propose an approach based on finite sample corrections and wild bootstrap to testing for $H_0 : \gamma_i = 0$ in a standard panel factor model defined as $y_{it} = \gamma_i' f_t + \epsilon_{it}$.

From a methodological point of view, we use statistics based on extrema of the estimated $\gamma_i$ and $f_t$, in a similar fashion to the tests for slope homogeneity developed by Kapetanios (2003) and Westerlund and Hess (2011). From a technical point of view, in our proofs we use similar arguments to the changepoint literature (see e.g. Csörgö and Hórvath, 1997): we approximate the sequences of estimated parameters with sequences of normals, and apply Extreme Value Theory (EVT henceforth). In this respect, our paper is a first attempt to systematize the use of extrema of estimated parameters in the context of a panel regression with unobservable common factors. As far as small sample properties are concerned, we show through a Monte Carlo exercise that the tests have correct size and satisfactory power for different levels of the signal-to-noise ratio and for several simulation designs.

The paper is organized as follows. The estimation procedure, and the asymptotics of the estimates of $\gamma_i$ and $f_t$ are in Section 2; Section 3 contains results about the two tests mentioned above. Section 4 discusses alternative testing approaches. Section 5 contains a validation of our theory through synthetic data. Section 6 concludes.

NOTATION. We use "$\longrightarrow$" to denote the ordinary limit; "$\xrightarrow{d}$" and "$\xrightarrow{p}$" to denote convergence in distribution and in probability respectively; and we use "a.s." as short-hand for "almost surely". The Frobenius norm of a matrix $A$ is denoted as $\|A\| = \sqrt{tr\,(A'A)}$, where $tr\,(A)$ denotes the trace of $A$. Definitional equality is denoted as "$\equiv$". Other notation is defined throughout the paper and in Appendix.

## 2 Estimation

In model (1)-(2), where $x_{it}$ is $m$-dimensional and $f_t$ is $r$-dimensional, we consider the following notation, which we use throughout the whole paper. We define $F = (f_1, ..., f_T)'$; $X_i = (x_{i1}, ..., x_{iT})'$; $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iT})'$; $y_i = (y_{i1}, ..., y_{iT})'$; $z_{it} = (y_{it}, x_{it}')'$; $z_i = (z_{i1}, ..., z_{iT})'$ and $\bar{H}_w = n^{-1} \sum_{i=1}^{n} z_i$. We also define the matrices $\bar{M}_w = I_T - \bar{H}_w \left(\bar{H}_w' \bar{H}_w\right)^{-1} \bar{H}_w'$ and

$$C_i = [\gamma_i | \Lambda_i'] \begin{bmatrix} 1 & 0_{1\times m} \\ \beta_i & I_m \end{bmatrix},$$

for each $i$. Based on this, the $\beta_i$s in (1) can be estimated as

$$\tilde{\beta}_i = \left(\frac{X_i' \bar{M}_w X_i}{T}\right)^{-1} \left(\frac{X_i' \bar{M}_w y_i}{T}\right), \tag{3}$$

which is the CCE estimator of Pesaran (2006); it holds that $\tilde{\beta}_i - \beta_i = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{1}{n}\right)$.

In order to estimate $\gamma_i$ and $f_t$, we propose the following two-step procedure.

**Step 1** Estimate the $\beta_i$s using the CCE estimator, and compute the residuals $\tilde{v}_i = y_i - X_i \tilde{\beta}_i$.

**Step 2** Apply the PC estimator to $\tilde{v}_i$, obtaining $\hat{\gamma}_i$ and $\hat{f}_t$ under the restrictions $\hat{F}'\hat{F} = TI_r$ and $n^{-1}\sum_{i=1}^n \hat{\gamma}_i \hat{\gamma}_i'$ diagonal.

In Step 2, $\hat{F}$ is calculated as $\sqrt{T}$ times the $r$ largest eigenvectors of $\frac{1}{nT}\sum_{i=1}^n \tilde{v}_i \tilde{v}_i'$. Similarly, $\hat{\gamma}_i$ is computed as

$$\hat{\gamma}_i = \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} \left(\hat{F}' M_{Xi} y_i\right), \tag{4}$$

with $M_{Xi} = I_T - X_i \left(X_i' X_i\right)^{-1} X_i'$. In (1), $\gamma_i$ and $f_t$ are not separately identifiable; as is typical in this literature, we only manage to estimate a rotation of $\gamma_i$ and $f_t$, say $H^{-1}\gamma_i$ and $H' f_t$. However, for our purposes knowing $H^{-1}\gamma_i$ and $H' f_t$ is as good as knowing $\gamma_i$ and $f_t$. We point out that the results in this paper do not strictly require the CCE estimator in Step 1: our results keep holding as long as the $\beta_i$s are estimated at a rate $O_p\left[\min\left\{T^{-1/2}, n^{-1}\right\}\right]$. Thus, the CCE is only a possible choice. Alternatives, like the Song (2013) estimator, which extends Bai (2009a) IE estimator to the case of heterogeneous slopes, may be used instead. The Song (2013) estimator obtains the same rate of convergence as for the CCE estimates of the individual slopes. In the remainder of the paper, we show our results based on employing the CCE in Step 1.

Consider the following assumptions.

**Assumption 1**. *[error terms: serial and cross sectional dependence] (i)* $E\left(\epsilon_{it}\right) = 0$ and $E\left|\epsilon_{it}\right|^{12} < \infty$; *(ii)* (a) $\sum_{t=1}^T \left|E\left(\epsilon_{it}\epsilon_{is}\right)\right| \leq M$ for all $i$ and $s$, (b) $\sum_{i=1}^n \sum_{j=1}^n \left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| \leq Mn$ for all $t$ and $s$, (c)

$\sum_{t=1}^{T}\sum_{s=1}^{T}|E(\epsilon_{it}\epsilon_{is})| \leq MT$ for all $i$, (d) $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}|E(\epsilon_{it}\epsilon_{js})| \leq M(nT)$; *(iii)* (a) $E\left|(nT)^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\epsilon_{it}\right|^{2} \leq M$, (b) $\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{v=1}^{T}\sum_{u=1}^{T}|E(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv})| \leq MT^{2}$, (c) $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{u=1}^{T}|E(\epsilon_{it}\epsilon_{is}\epsilon_{ju}\epsilon_{js})| \leq M(nT)$ for all $u$, (d) $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}|E(\epsilon_{it}\epsilon_{kt}\epsilon_{js}\epsilon_{ks})| \leq M(nT)$ for all $k$; *(iv)* (a) $E\left|\sum_{t=1}^{T}\epsilon_{it}\right|^{r} \leq ME\left|\sum_{t=1}^{T}\epsilon_{it}^{2}\right|^{r/2}$ for all $i$, $r<12$, (b) $E\left|\sum_{i=1}^{n}\epsilon_{it}\right|^{r} \leq ME\left|\sum_{i=1}^{n}\epsilon_{it}^{2}\right|^{r/2}$ for all $t$, $r<12$.

**Assumption 2**. *[regressors and common factors] (i)* $E\|\epsilon_{it}^{x}\|^{12}<\infty$ and $E\|f_t\|^{12}<\infty$; *(ii)* $T^{-1}\sum_{t=1}^{T}f_tf_t' \xrightarrow{p} \Sigma_f$ as $T\to\infty$ with $\Sigma_f$ non-singular; *(iii)* $\{\epsilon_{it}^{x}, f_t\}$ and $\{\epsilon_{js}\}$ are mutually independent for all $i$, $j$, $t$, $s$; *(iv)* $E\left|\sum_{t=1}^{T}x_{it}\epsilon_{it}\right|^{r} \leq ME\left|\sum_{t=1}^{T}(x_{it}\epsilon_{it})^{2}\right|^{r/2}$ for all $i$, $r\leq 6$.

**Assumption 3**. *[slopes and loadings] (i)* $\{\beta_i\}$ is independent of $\{\epsilon_{jt}, \epsilon_{jt}^{x}, f_t\}$ for all $i$, $j$, $t$; *(ii)* $E\|\beta_i\|^{2+\delta}<\infty$ for some $\delta>0$; *(iii)* the $\gamma_i$s are non stochastic and such that $\max_i\|\gamma_i\|<\infty$ and $n^{-1}\sum_{i=1}^{n}\gamma_i\gamma_i' \to \Sigma_\gamma$ as $n\to\infty$ with $\Sigma_\gamma$ non-singular.

**Assumption 4**. *[Step 1 estimation] (i)* $l_{\min}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)>0$; $l_{\min}\left(\frac{X_i'M_FX_i}{T}\right)>0$ and $l_{\min}\left(\frac{F'M_{Xi}F}{T}\right)>0$ a.s. for all $i$, where $l_{\min}(\cdot)$ denotes the smallest eigenvalue; *(ii)* $C\equiv n^{-1}\sum_{i=1}^{n}C_i$ has rank $r\leq m+1$.

**Assumption 5.** *[Central Limit Theorems] (i)* (a) there exists a nonrandom, positive definite matrix $\Sigma_{fM,i}$ such that $p\lim_{T\to\infty}T^{-1}F'H'M_{Xi}HF=\Sigma_{fM,i}$, (b) $T^{-1/2}F'H'M_{xi}\epsilon_i \xrightarrow{d} N(0,\Sigma_{fMe,i})$, where $\Sigma_{fMe,i}=p\lim_{T\to\infty}T^{-1}F'H'M_{xi}\epsilon_i\epsilon_i'M_{xi}HF$, for all $i$; *(ii)* $n^{-1/2}\sum_{i=1}^{n}\gamma_i\epsilon_{it} \xrightarrow{d} N(0,\Phi_{\gamma\epsilon,t})$, where $\Phi_{\gamma\epsilon,t}=p\lim_{n\to\infty}n^{-1}\gamma_i\gamma_i'\epsilon_{it}\epsilon_{it}$, for all $t$.

Broadly speaking, Assumptions 1-4 are needed to prove the consistency of the estimated common factors and loadings. Assumption 4 is specific to the CCE estimator, employed in Step 1. Assumption 5 is required when deriving the asymptotic distributions.

In particular, Assumption 1 deals with the error term $\epsilon_{it}$, and it allows for serial and cross dependence. The conditions in parts *(ii)* and *(iii)* of the assumption resemble closely (and in some cases are exactly the same as) those in Bai (2003) and Bai (2009a), and can be shown immediately if $\epsilon_{it}$ is assumed to be independent. Part *(i)* requires the existence of the 12-th moment of $\epsilon_{it}$, which is stronger than what the literature normally considers - e.g. in Bai (2009a), assuming $E|\epsilon_{it}|^{8}<\infty$ suffices. In our context, the existence of the 12-th moment is needed in order to derive consistency of $\hat{\gamma}_i$ and $\hat{f}_t$ (see in particular the proof of Lemma A.1). Finally, part *(iv)* contains Burkholder-type inequalities: these could be shown directly under more specific assumptions on the degree of serial and cross sectional dependence. For example, part (a) holds immediately if one assumes that $\epsilon_{it}$ is a Martingale Difference Sequence (MDS) across $t$ (the same holds for part (b), under the MDS assumption across $i$) - see e.g. Lin and Bai (2010, p.108).

As far as Assumption 2 is concerned, we allow for serial and cross sectional dependence in both the $\epsilon_{it}^{x}$s and in the common factors $f_t$. The requirement in part *(ii)* is standard in the literature (see e.g. Assumption B in Bai, 2009a), and it entails that common factors are "strong" in the sense of Chudik, Pesaran and Tosetti (2011) (see in particular Assumption 3). Finally, according to part *(iii)*, the $x_{it}$s are strictly exogenous. Assumption 3 is standard. Assumption 4 is specific to the CCE estimator of the $\beta_i$s, employed in Step 1. Particularly, the rank condition in part *(ii)* is the same as equation (21) in Pesaran (2006), and it guarantees the consistency of the $\tilde{\beta}_i$s.

Finally, Assumption 5 contains two CLT-type results which are employed when deriving the limiting distributions of the estimated common factors and loadings: parts *(i)* and *(ii)* can be compared with Assumption F in Bai (2003).

We now turn to studying the asymptotics of $\hat{\gamma}_i$ and $\hat{f}_t$.

**Theorem 1** *Let Assumptions 1-4 hold; then, for every $i$*

$$\hat{\gamma}_i - H^{-1}\gamma_i = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{n}\right). \tag{5}$$

*Let Assumptions 1-5 hold. As $(n, T) \to \infty$ with $\frac{\sqrt{T}}{n} \to 0$*

$$\sqrt{T}\left(\hat{\gamma}_i - H^{-1}\gamma_i\right) \xrightarrow{d} N\left(0, \Sigma_{\gamma i}\right), \tag{6}$$

*where $\Sigma_{\gamma i} = \Sigma_{fM,i}^{-1}\Sigma_{fMe,i}\Sigma_{fM,i}^{-1}$ and $\Sigma_{fM,i}$ and $\Sigma_{fMe,i}$ are the probability limits of $T^{-1}\left(F'H'M_{Xi}HF\right)$ and $T^{-1}\left(F'H'M_{Xi}\epsilon_i\epsilon_i'M_{Xi}HF\right)$, respectively.*

Theorem 1 can be compared with Theorem 2 in Bai (2003, p.147): the rates of convergence in (5) are exactly the same. On the other hand, the limiting distribution of $\sqrt{T}\left(\hat{\gamma}_i - H^{-1}\gamma_i\right)$ in (6) is different from the one in Theorem 2 in Bai (2003): this is due to the presence, in our context, of the idiosyncratic regressors $x_{it}$.

We use the estimator of $\Sigma_{\gamma i}$ proposed in (Bai, 2003, p.150)

$$\hat{\Sigma}_{\gamma_i} = \left(Q_i'\right)^{-1}\Phi_i\left(Q_i\right)^{-1} \tag{7}$$

where $Q_i = T^{-1}(\hat{F}'M_{Xi}\hat{F})$, and $\Phi_i = D_{0,i} + \sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\left(D_{j,i} + D_{j,i}'\right)$, with $D_{j,i} = T^{-1}\sum_{t=j+1}^{T}$ $\widehat{fx}_t'\ \widehat{fx}_{t-j}\ \hat{\epsilon}_{it}\ \hat{\epsilon}_{it-j}$, where $\widehat{fx}_t$ is the t-th row of $M_{Xi}\hat{F}$ and $\hat{\epsilon}_{it} = y_{it} - \hat{\beta}_i'x_{it} - \hat{\gamma}_i'\hat{f}_t$. The bandwidth $q$ is chosen so that $q \to \infty$ with $q/T^{1/4} \to 0$.

We now present the asymptotic results for $\hat{f}_t$.

**Theorem 2** *Let Assumptions 1-4 hold; then, for every t*

$$\hat{f}_t - H' f_t = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T}\right). \tag{8}$$

*Let Assumptions 1-5 hold. As* $(n, T) \to \infty$ *with* $\frac{\sqrt{n}}{T} \to 0$

$$\sqrt{n}\left(\hat{f}_t - H' f_t\right) \xrightarrow{d} N\left(0, \Sigma_{ft}\right), \tag{9}$$

*where* $\Sigma_{ft} = H\Sigma_f \Sigma_{\Gamma\epsilon,t}\Sigma_f H'$ *and* $\Sigma_{\Gamma\epsilon,t} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_i \gamma_j' \epsilon_{it}\epsilon_{jt}$.

Theorem 2 is the counterpart to Theorem 1 in Bai (2003, p.145). Rates of convergence and limiting distribution are exactly the same: the presence of individual specific regressors does not affect inference on the common factors.

By virtue of Theorem 2, the asymptotic covariance matrix of $\sqrt{n}\left(\hat{f}_t - H' f_t\right)$ can be estimated using equation (7) in Bai (2003, p.150). Specifically, letting $\hat{\epsilon} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_n)'$ with $\hat{\epsilon}_i = [\hat{\epsilon}_{i1}, ..., \hat{\epsilon}_{iT}]'$, and defining $V_{nT}$ as a diagonal matrix containing the $r$ largest eigenvalues of $\frac{1}{nT}\hat{\epsilon}\hat{\epsilon}'$ in descending order, the estimated $\Sigma_{ft}$ is

$$\hat{\Sigma}_{ft} = V_{nT}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}_i \hat{\gamma}_i' \hat{\epsilon}_{it}^2\right) V_{nT}^{-1}. \tag{10}$$

Note that $\Sigma_{\Gamma\epsilon,t}$ is estimated through $n^{-1}\sum_{i=1}^{n} \hat{\gamma}_i \hat{\gamma}_i' \hat{\epsilon}_{it}^2$, which is valid under cross sectional independence. It is not possible, in general, to estimate $\Sigma_{\Gamma\epsilon,t}$ consistently unless some ordering among the cross sectional units is assumed - see also Bai (2003, p.150).

Combining Theorems 1 and 2, we obtain the asymptotics for the estimated common component $c_{it} = \gamma_i' f_t$, defined as $\hat{c}_{it} = \hat{\gamma}_i' \hat{f}_t$.

**Corollary 1** *Let Assumptions 1-4 hold; then, for all i and t*

$$\hat{c}_{it} - c_{it} = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{11}$$

*Let Assumptions 1-5 hold. As* $(n, T) \to \infty$

$$\left(\frac{1}{n}\gamma_i' \Sigma_{ft}\gamma_i + \frac{1}{T} f_t' \Sigma_{\gamma i} f_t\right)^{-1/2} \left(\hat{c}_{it} - c_{it}\right) \xrightarrow{d} N(0, 1), \tag{12}$$

*where* $\Sigma_{ft}$ *is defined in Theorem 2 and* $\Sigma_{\gamma i}$ *in Theorem 1.*

After discussing the asymptotic properties of $\hat{\gamma}_i$ and $\hat{f}_t$, we turn to deriving tests for the null of no factor structure.

## 3 Testing for no factor structure

In this section, we discuss and compare two approaches to testing for the null of no factor structure in (1). Motivated by Sarafidis, Yamagata and Robertson (2009), we study tests for, respectively: *(a)* the null of cross-sectional homogeneity of the loadings $\gamma_i$s; and *(b)* the null of homogeneity, over time, of the $f_t$s.

Formally, we propose two tests for the null hypotheses:

$$H_0^a \quad : \quad \gamma_i = \gamma \text{ for all } i; \tag{13}$$

$$H_0^b \quad : \quad f_t = f \text{ for all } t. \tag{14}$$

Both (13) and (14) entail that there is no real factor structure in (1). Consider (13) first. When $H_0^a$ holds, equation (1) can be rewritten as

$$y_{it} = \varphi_t + \beta_i' x_{it} + \epsilon_{it}, \tag{15}$$

where we have defined $\varphi_t = \gamma' f_t$. Thus, under $H_0^a$, model (1) boils down to a standard panel specification with a time effect. Similarly, under $H_0^b$ in (14), equation (1) can be rewritten as

$$y_{it} = \varphi_i + \beta_i' x_{it} + \epsilon_{it}, \tag{16}$$

where we have defined $\varphi_i = \gamma_i' f$. Therefore, under $H_0^b$, model (1) is tantamount to a standard panel specification with a unit specific effect.

The considerations made above also entail that testing for (13) and (14) is equivalent to testing for strong cross dependence among the $y_{it}$s. Sarafidis, Yamagata and Robertson (2009) propose a test for cross dependence (albeit in a different context) based on verifying the null that loadings are homogeneous, i.e. $\gamma_i = \gamma$. Our paper extends the contribution by Sarafidis, Yamagata and Robertson (2009) to our context, and complements it by also considering a test for (14). A similar approach to testing for factor structures versus models with individual or time dummies is also suggested in Bai (2009a).

In order to test for (13) and (14), we propose two tests based directly on the results in Section 2, i.e. on the estimates of $\gamma_i$ and $f_t$. Specifically, we propose two max-type statistics, where the maximum is taken over the deviation of the individual estimate of $\gamma_i$ (resp. of $f_t$) with respect to their cross-sectional (resp. time) average. This approach has been proposed, in the context of testing for poolability with observable regressors, by Westerlund and Hess (2011), whose simulations show that the power properties are very promising, although issues may arise in presence of ties (Hall and Miller, 2010). In our context, we show that tests based on max-type statistics have power even versus alternatives whereby only one unit/time period has heterogeneous loadings/common factors. Other approaches to testing for $H_0^a$ and $H_0^b$ are discussed in Section 4.

Define $\widehat{\bar{\gamma}} = n^{-1} \sum_{i=1}^{n} \hat{\gamma}_i$ and $\widehat{\bar{f}} = T^{-1} \sum_{t=1}^{T} \hat{f}_t$. We propose the following max-type test statistics:

$$S_{\gamma,nT} \equiv \max_{1 \leq i \leq n} \left[ T \left( \hat{\gamma}_i - \widehat{\bar{\gamma}} \right)' \hat{\Sigma}_{\gamma i}^{-1} \left( \hat{\gamma}_i - \widehat{\bar{\gamma}} \right) \right], \tag{17}$$

$$S_{f,nT} \equiv \max_{1 \leq t \leq T} \left[ n \left( \hat{f}_t - \widehat{\bar{f}} \right)' \hat{\Sigma}_{ft}^{-1} \left( \hat{f}_t - \widehat{\bar{f}} \right) \right]. \tag{18}$$

We point out that under the null hypotheses $H_0^a$ and $H_0^b$, the spaces spanned by the loadings and by the factors (respectively) have rank equal to one. This fact was already noted by Sarafidis, Yamagata and Robertson (2009) who, building on it, suggest running their test setting $r = 1$. This can be applied to our context also: $S_{\gamma,nT}$ and $S_{f,nT}$ can be used setting $r = 1$, which avoids having to estimate $r$.

From a methodological perspective, this entails that tests based on (17) and (18) can be implemented without prior knowledge of the number of factors: thus, testing does not require estimation of $r$ as a preliminary step. Indeed, we note that tests for (17) and (18) are to be implemented *before* determining $r$. If the null is not rejected, the conclusion can be drawn that no factor structure is needed, and either (15) or (16) is the correct specification. Conversely, if the null is rejected, then it follows that there is a genuine factor structure. Hence, the next step is determining the number of latent common factors $r$, e.g. by applying some information criteria as discussed in Bai and Ng (2002) and Bai (2009b). The asymptotic properties of the estimated common factors, loadings and common components are those given in Section 2.

We now report the results on tests based on $S_{\gamma,nT}$ (Theorem 3) and on $S_{f,nT}$ (Theorem 4). For both test statistics, a heuristic preview of the main arguments used in the proofs of both theorems is as follows. Referring to (17) as a benchmark example, we approximate the sequence of the estimation errors $\sqrt{T} \left( \hat{\gamma}_i - H^{-1} \gamma_i \right)$ with a sequence of normally distributed random variables, plus an error term whose supremum taken over $n$ is negligible. In light of this, the proofs are similar, in spirit, to the ones found in the changepoint literature (see e.g. Csörgö and Hórvath, 1997).

### 3.1 Testing for $H_0^a : \gamma_i = \gamma$

In this section we report the asymptotics of $S_{\gamma,nT}$ under the null $H_0^a$, and we analyse the consistency of tests based on $S_{\gamma,nT}$. We show that, as $(n,T) \to \infty$ under some restrictions on the relative speed of divergence, $S_{\gamma,nT}$ (suitably normalised) converges to a Gumbel distribution. Further, we also show that tests based on $S_{\gamma,nT}$ have nontrivial power versus alternative hypotheses shrinking at a rate $O_p\left(\sqrt{\frac{\ln n}{T}}\right)$.

Let $k_1$ be the largest number for which $E\left|\epsilon_{it}\right|^{k_1}$, $E\left\|x_{it}\right\|^{k_1}$ and $E\left\|f_t\right\|^{k_1}$ are finite. In view of Assumption 1, $k_1 \geq 12$. Consider the following assumptions, which complement Assumptions 1 and 2, imposing further conditions on the form of time and cross sectional dependence.

**Assumption 6**. *[serial dependence]* Let $\delta > 0$ and $\alpha \in (1, +\infty)$: *(i)* $\epsilon_{it}$, $f_t$ and $x_{it}$ are $L_{2+\delta}$-NED (Near Epoch Dependent) of size $\alpha$ on a uniform mixing base $\{v_t\}_{t=-\infty}^{+\infty}$ of size $-r/(r-2)$ and $r > \frac{2\alpha-1}{\alpha-1}$; *(ii)* (a) letting $V_{iT}^{f\epsilon} \equiv T^{-1}\ E\left[\left(\sum_{t=1}^T f_t\epsilon_{it}\right)\left(\sum_{t=1}^T f_t\epsilon_{it}\right)'\right]$, $V_{iT}^{f\epsilon}$ is positive definite uniformly in $T$, and as $T \to \infty$, $V_{iT}^{f\epsilon} \to V_i^{f\epsilon}$ with $\left\|V_i^{f\epsilon}\right\| < \infty$, (b) the same holds for $V_{iT}^{x\epsilon} \equiv T^{-1}\ E\left[\left(\sum_{t=1}^T x_{it}\epsilon_{it}\right)\left(\sum_{t=1}^T x_{it}\epsilon_{it}\right)'\right]$, $V_{iT}^{fx} \equiv T^{-1}E\left(\bar{w}_{iT}^{fx}\bar{w}_{iT}^{fx\prime}\right)$ with $\bar{w}_{iT}^{fx} = vec\left(\sum_{t=1}^T f_t x_{it}'\right) - E\left[vec\left(\sum_{t=1}^T f_t x_{it}'\right)\right]$, and $V_{iT}^{xx} = T^{-1}E\left(\bar{w}_{iT}^{xx}\bar{w}_{iT}^{xx\prime}\right)$ with $\bar{w}_{iT}^{xx} = vec\left(\sum_{t=1}^T x_{it}x_{it}'\right) - E\left[vec\left(\sum_{t=1}^T x_{it}x_{it}'\right)\right]$; *(iii)* (a) letting $w_{kt}^{f\epsilon}$ be the $k$-th element of $f_t\epsilon_{it}$ and defining $S_{kT,m}^{f\epsilon} \equiv \sum_{t=m+1}^{m+T} w_{kt}^{f\epsilon}$, there exists a positive definite matrix $\bar{\Omega}^{f\epsilon} = \left\{\varpi_{kh}^{f\epsilon}\right\}$ such that $T^{-1}\left|E\left[S_{kT,m}^{f\epsilon}S_{hT,m}^{f\epsilon}\right] - \varpi_{kh}^{f\epsilon}\right| \leq MT^{-\psi}$, for all $k$ and $h$ and uniformly in $m$, with $\psi > 0$, (b) the same holds for $x_{it}\epsilon_{it}$.

**Assumption 7**. *[cross sectional dependence]* It holds that $T^{-1}\sum_{t=1}^T\sum_{s=1}^T\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right|\ln n \to 0$ as $(n,T) \to \infty$ for all $i \neq j$.

Assumptions 6 and 7 complement Assumptions 1 and 2, by adding further requirements on the form of serial dependence and on the amount of cross dependence respectively.

More specifically, Assumption 6 specifies the amount of memory allowed in the series $\epsilon_{it}$, $f_t$ and $x_{it}$ - these all have, by Assumptions 1 and 2, finite moments up to order 12. The assumption is needed in order to prove an a.s. version of the Invariance Principle (IP), and it is a quite general specification for the form and amount of serial dependence. Part *(iii)* is a bound on the growth rate of the variance of partial sums, and it is the same as equation (1.5) in Eberlein (1986); see also Assumption A.3 in Corradi (1999).

As far as Assumption 7 is concerned, it complements the summability conditions in Assumption 1 by allowing for some cross dependence. In essence, it requires that $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right|$ declines (faster than $\ln n$) as $n$ passes to infinity. This assumption is similar to the so-called "Berman condition" (Berman, 1964), which is employed in EVT for dependent time series data; we refer to Assumption 9 below for further explanations on how the Berman condition works in the case of time series data. By way of comparison, Assumption 7 can be viewed as a complement to Assumption 1*(ii)*(d), since it contains the same summation across $t$. As far as the amount of cross sectional dependence is concerned, the assumption is quite weak; as an example, it would be satisfied if $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| = o\left(\ln^{-1} n\right)$ for all $i \neq j$, which is a much weaker requirement than the one in Assumption 1*(ii)*(d).

Let the critical value $c_{\alpha,n}$ be defined such that $P\left(S_{\gamma,nT} \leq c_{\alpha,n}\right) = 1-\alpha$ under $H_0^a$, and let $\Gamma\left(\cdot\right)$ denote the Gamma function. It holds that:

**Theorem 3** *Let Assumptions 1-4 and 6-7 hold, and let* $(n,T) \to \infty$ *with*

$$\frac{\sqrt{T}n^{2/k_1}}{n} + \frac{n^{4/k_1}}{T} \to 0. \tag{19}$$

*Under* $H_0^a$*, it holds that*

$$P\left(A_n S_{\gamma,nT} \leq x + B_n\right) = e^{-e^{-x}}, \tag{20}$$

*where* $A_n = \frac{1}{2}$ *and* $B_n = \ln\left(n\right) + \left(\frac{r}{2}-1\right)\ln\ln\left(n\right) - \ln\Gamma\left(\frac{r}{2}\right)$*. Under the alternative* $H_1^a : \gamma_i = \gamma + c_i$ *for at least one* $i$*, if*

$$\frac{T}{\ln n}\left\|c_i\right\|^2 \to \infty, \tag{21}$$

*it holds that* $P\left(S_{\gamma,nT} > c_{\alpha,n}\right) = 1$.

Theorem 3 states that $S_{\gamma,nT}$ has a Gumbel distribution. This holds in the joint limit $(n,T) \to \infty$, with the restrictions specified in (19). Since $k_1 \geq 12$, the latter condition requires $\frac{T}{n^{5/3}} \to 0$, which is marginally stricter than the condition $\frac{\sqrt{T}}{n} \to 0$ needed in for (6). Also, (19) needs that $\frac{n^{4/k_1}}{T} \to 0$; this becomes, under Assumptions 1*(i)* and 2*(i)*, $\frac{n}{T^3} \to 0$. It is interesting to note that, based on equation (45) in Appendix B, if all moments exist (as is the case with Gaussian variables), then (19) reduces to $\frac{\sqrt{T}\ln n}{n} + \frac{\ln^4 n}{T} \to 0$, which is essentially the same as in Theorem 1.

Equation (20) also provides a rule to calculate asymptotic critical values $c_{\alpha,n}$, which are given by

$$c_{\alpha,n} = 2B_n - \ln\left|\ln\left(1-\alpha\right)\right|^2. \tag{22}$$

Thus, for a given level $\alpha$, $c_{\alpha,n}$ is nuisance free, and it depends only on the cross-sectional sample size, $n$. A well known issue in EVT is that convergence to Extreme Value distributions is in general rather slow. Canto e Castro (1987) shows that the rate of convergence for the maximum of a sequence of random variables following a Gamma distribution is $O\left(1/\ln^2 n\right)$. Unreported Monte Carlo evidence shows that tests based on using $c_{\alpha,n}$ perform quite well, although they are a bit oversized. As an alternative, one can replace $B_n$ with $F_{\chi_r}^{-1}(1-1/n)$, where $F_{\chi_r}^{-1}(\cdot)$ is the inverse of the cumulative distribution function of a chi-square with $r$ degrees of freedom, see Embrechts, Klüppelberg and Mikosch (1997).

As far as consistency of the test is concerned, equation (21) shows that nontrivial power is attained versus local alternatives shrinking at a rate $O_p\left(\sqrt{\frac{\ln n}{T}}\right)$. Thus, when using max-type statistics such as $S_{\gamma,nT}$, $n$ does not play a role in enhancing the power of the test. On the other hand, the test is powerful as long as just one $\gamma_i$ is different from the others.

### 3.2 Testing for $H_0^b : f_t = f$

We report the asymptotics of $S_{f,nT}$ under $H_0^b$, and its consistency. Similarly to the previous subsection, we show that, as $(n,T)\to\infty$ under some restrictions on the relative speed of divergence, $S_{f,nT}$ (suitably normalised) converges to a Gumbel distribution. Further, we also show that tests based on $S_{f,nT}$ have nontrivial power versus alternative shrinking at a rate $O_p\left(\sqrt{\frac{\ln T}{n}}\right)$.

Let $k_2$ be the largest number such that $E\left\|f_t\right\|^{k_2}$, $E\left\|x_{it}\right\|^{k_2}$ and $E\left|\epsilon_{it}\right|^{k_2}$ are all finite. In view of Assumptions 1 and 2, $k_2 \geq 12$. Consider also the following assumption, which, as in the previous section, complement Assumptions 1 and 2 by adding further structure to the serial and cross sectional dependence of the series.

**Assumption 8**. *[cross sectional dependence]* Let $\delta > 0$ and $\alpha \in (1,+\infty)$: *(i)* $\epsilon_{it}$ is $L_{2+\delta}$-NED across $i$, of size $\alpha$ on a uniform mixing base $\{v_i\}_{i=-\infty}^{+\infty}$ of size $-r/(r-2)$ and $r > \frac{2\alpha-1}{\alpha-1}$; *(ii)* letting $V_{tn}^{\epsilon\epsilon} = n^{-1}\ E\left[\left(\sum_{i=1}^n \epsilon_{it}\right)\left(\sum_{i=1}^n \epsilon_{it}\right)\right]$, $V_{tn}^{\epsilon\epsilon}$ is positive definite uniformly in $n$, and as $n\to\infty$, $V_{tn}^{\epsilon\epsilon}\to V_t^{\epsilon\epsilon}$ with $\|V_t^{\epsilon\epsilon}\| < \infty$; *(iii)* letting $S_{mt}^{\epsilon} = \sum_{i=m+1}^{m+n}\epsilon_{it}$ there exists a positive constant $\varpi^{\epsilon\epsilon}$ such that $n^{-1}\left|E\left(S_{mt}^{\epsilon 2}\right)-\varpi^{\epsilon\epsilon}\right| \leq Mn^{-\psi''}$ uniformly in $m$, with $\psi''>0$.

**Assumption 9**. *[serial dependence]* It holds that $\lim_{k\to\infty} n^{-1}\sum_{i=1}^n\sum_{j=1}^n \left|E\left(\epsilon_{it}\epsilon_{jt-k}\right)\right|\ln k = 0$ as $(n,T)\to\infty$.

Assumption 8 is very similar, in spirit, to Assumption 6, and it requires that $\epsilon_{it}$ is NED across $i$. By virtue of Assumption 8, an a.s. IP holds for $\sum_{i=1}^n \epsilon_{it}$ and for $\sum_{i=1}^n \epsilon_{it}^2$. The definition of NED for spatial processes has been studied in Jenish and Prucha (2012), and we refer to that paper for details.

Assumption 9 is the so-called "Berman condition" (Berman, 1964): as mentioned when discussing Assumption 7, standard EVT, which holds for *i.i.d.* data, can be applied under such condition, yielding the same results as in the case of independence. Berman condition holds as long as serial correlations have at least a logarithmic rate of decay, and it is a sufficient condition used to verify more general mixing conditions which are typical of EVT (and more difficult to verify; see e.g. Leadbetter and Rootzen, 1988). Assumption 9 is a very mild requirement: for example in the case of ARMA processes, typically the autocovariances have an exponential rate of decay (see e.g. Hannan and Kavalieris, 1986), which is more than enough to ensure that Assumption 9 holds. Further, Assumption 9 can be shown to hold in contexts where the autocorrelation function is not absolutely summable, as e.g. fractional ARIMA processes. In our context, Assumption 9 can be compared to Assumption 1*(ii)*(d), and it contains the same summation across $i$.

Let the critical value $c_{\alpha,T}$ be defined such that $P\left(S_{f,nT} \leq c_{\alpha,T}\right) = 1-\alpha$ under $H_0^b$. It holds that:

**Theorem 4** *Let Assumptions 1-4 hold and 8-9, and let $(n,T) \to \infty$ with*

$$\frac{\sqrt{n}T^{1/k_2}}{T} + \frac{T^{4/k_2}}{n} \to 0. \tag{23}$$

*Under $H_0^b$, it holds that*

$$P\left[A_T S_{f,nT} \leq x + B_T\right] = e^{-e^{-x}}, \tag{24}$$

*where $A_T = \frac{1}{2}$ and $B_T = \ln(T) + \left(\frac{r}{2}-1\right)\ln\ln(T) - \ln\Gamma\left(\frac{r}{2}\right)$. Under the alternative $H_1^b : f_t = f + c_t$ for at least one $t$, if*

$$\frac{n}{\ln T}\left\|c_t\right\|^2 \to \infty, \tag{25}$$

*it holds that $P\left(S_{f,nT} > c_{\alpha,T}\right) = 1$.*

Theorem 4 is very similar to Theorem 3; convergence to the Gumbel distribution under the null is shown for $(n,T) \to \infty$ jointly under some restrictions between $n$ and $T$, spelt out in (23). Specifically, it is required that $\frac{T^{1/k_2}\sqrt{n}}{T} \to 0$; since $k_2 \geq 12$, the former restriction is, at most, $\frac{n}{T^{11/6}} \to 0$. This is only marginally stronger than $\frac{\sqrt{n}}{T} \to 0$, which is required for (9) to hold. Similarly, requiring that $\frac{T^{4/k_2}}{n} \to 0$ entails $\frac{T}{n^3} \to 0$. As in the case of Theorem 3, the test should be applied when $n$ is not exceedingly larger than $T$, and vice versa. Using (45), under the assumptions that all moments exist, (23) becomes $\frac{\sqrt{n}\ln T}{T} + \frac{\ln^4 T}{n} \to 0$ - again very close to the restriction needed in Theorem 2.

Critical values for a test of level $\alpha$ can be calculated as

$$c_{\alpha,T} = 2B_T - \ln\left|\ln(1-\alpha)\right|^2; \tag{26}$$

alternatively, $B_T$ can be approximated by $F^{-1}_{\chi_r}(1-1/T)$.

As far as power is concerned, (25) stipulates that the test is consistent versus alternatives shrinking as $O\left(\sqrt{\frac{\ln T}{n}}\right)$. Similarly to Theorem 3, it suffices that $f_t$ differs from $f$ in just one period $t$ for the test to reject $H_0^b$.

# 4 Discussion - other testing approaches

This section discusses other possible approaches to test for (13) and (14). We show that it is in general not possible to use average-type statistics of the estimated $\gamma_i$ and $f_t$ (Section 4.1). We also discuss tests based on applying the Hausman principle to the estimated slopes (Section 4.2).

## 4.1 Tests based on average-type statistics

Pesaran and Yamagata (2008) suggest using averages of $F$-statistics in order to test for the null of slope homogeneity in a model with observable regressors, viz.

$$\tilde{S}_{\gamma,nT} = \sqrt{\frac{n}{2r}}\frac{1}{n}\sum_{i=1}^{n}\left[T\left(\hat{\gamma}_i-\widehat{\bar{\gamma}}\right)'\hat{\Sigma}^{-1}_{\gamma i}\left(\hat{\gamma}_i-\widehat{\bar{\gamma}}\right)-r\right], \tag{27}$$

$$\tilde{S}_{f,nT} = \sqrt{\frac{T}{2r}}\frac{1}{T}\sum_{t=1}^{T}\left[n\left(\hat{f}_t-\widehat{\bar{f}}\right)'\hat{\Sigma}^{-1}_{ft}\left(\hat{f}_t-\widehat{\bar{f}}\right)-r\right]. \tag{28}$$

Similarly to the max-type statistics defined in (17) and (18), estimation of $r$ is not required, and tests can be carried out setting $r=1$.

We show that $\tilde{S}_{\gamma,nT}$ and $\tilde{S}_{f,nT}$ cannot be employed in our context: in essence, this is because $\tilde{S}_{\gamma,nT}$ and $\tilde{S}_{f,nT}$ diverge under the null as $(n,T)\to\infty$, so that tests based on (27) and (28) always reject the null of no factor structure.

Results are summarized in the following Theorem:

**Theorem 5** *Let Assumptions 1-4 hold.*

1. *If, in addition, as* $(n,T)\to\infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\epsilon_i' M_{xi}F\Sigma^{-1}_{fMe,i}F'M_{Xi}\epsilon_i - r\right] = O_p(1), \tag{29}$$

*then, under $H_0^a$ it holds that $\tilde{S}_{\gamma,nT} = O_p(1) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p\left(\sqrt{\frac{n}{T}}\right)$.*

2. *If, in addition, as $(n,T) \to \infty$*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\left(\frac{\hat{F}'F}{T}\right)\left(\frac{F'\hat{F}}{T}\right)\gamma_j\epsilon_{it}\epsilon_{jt} - r\right] = O_p(1), \tag{30}$$

*then, under $H_0^b$ it holds that $\tilde{S}_{f,nT} = O_p(1) + O_p(\sqrt{n}) + O_p\left(\frac{n}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{T}}{n}\right)$.*

Theorem 5 shows that, under the respective null hypotheses, both average-type statistics diverge, and therefore cannot be employed.

## 4.2 Tests based on the Hausman principle

Building on Bai (2009a, Section 9), tests could be constructed indirectly using a pooled estimator of the $\beta_i$s.[1]

In order to illustrate the idea, define the average slope $\beta = E(\beta_i)$. Estimation of $\beta$ could be based on pooling the estimates of the individual $\beta_i$s:

$$\hat{\beta}^{CCE/IE} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\beta}_i^{CCE/IE}.$$

We use the notation $\hat{\beta}^{CCE}$ and $\hat{\beta}^{IE}$ according as the $\tilde{\beta}_i$s are computed using the individual CCE estimators (Pesaran, 2006) or the individual IE estimators (see Song, 2013) respectively. One can expect that under either null $H_0^a$ and $H_0^b$, both the CCE and the IE estimators are consistent, since no assumption for the consistency of either estimator is violated. The Hausman principle can therefore be applied upon finding another estimator which is consistent, and more efficient, under the null - Bai (2009a) points out that, in the context of slope homogeneity, estimators based on the "between" and "within" transformation should be more efficient under the null.

*Testing for $H_0^a : \gamma_i = \gamma$*

Under $H_0^a$, an alternative estimator for $\beta$ is

$$\hat{\beta}^{bw} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{y}_{it}\right),$$

[1] We wish to thank the anonymous Associate Editor for asking the question that led to the results in this Section.

with $\dot{x}_{it} = x_{it} - n^{-1}\sum_{i=1}^{n} x_{it}$ and $\dot{y}_{it} = y_{it} - n^{-1}\sum_{i=1}^{n} y_{it}$; this is the Mean-Group version of the "between" estimator, as also suggested in Bai (2009a). It can be expected that, under $H_0^a$, $\hat{\beta}^{bw}$ is consistent and should be more efficient than $\hat{\beta}^{CCE}$ and $\hat{\beta}^{IE}$. Hence, tests for $H_0^a$ could be based on

$$S_{\gamma,nT}^{IE/CCE} = n\left(\hat{\beta}^{IE/CCE} - \hat{\beta}^{bw}\right)'\left[Var\left(\hat{\beta}^{IE/CCE} - \hat{\beta}^{bw}\right)\right]^{-1}\left(\hat{\beta}^{IE/CCE} - \hat{\beta}^{bw}\right).$$

Let the critical value $c_{\alpha,n}$ be defined such that $P\left(S_{\gamma,nT}^{IE/CCE} \leq c_{\alpha,n}\right) = 1-\alpha$ under $H_0^a$. It holds that:

**Theorem 6** *Let Assumptions 1, 2, 3(i)-(ii) and 4 hold. As $(n,T)\to\infty$ with $\frac{\sqrt{n}}{T}\to 0$, under $H_0^a$, $S_{\gamma,nT}^{IE} \xrightarrow{d} \chi_m^2$. Assume further that $\gamma_i$ is i.i.d. (and independent of all other quantities) with mean $\gamma$ and $E\left\|\gamma_i\right\|^{2+\delta} < \infty$. Then, under the alternative $H_1^a : \gamma_i \neq \gamma_j$ for $i \neq j$, as $(n,T)\to\infty$ it holds that $P\left(S_{\gamma,nT}^{IE} > c_{\alpha,n}\right) < 1$. The same results holds for $S_{\gamma,nT}^{CCE}$ as $\min\{n,T\}\to\infty$.*

Theorem 6 is, in essence, a negative result. It is possible to construct a test statistic that does not diverge under the null, and which has a "standard" limiting distribution - this can be contrasted with Theorem 7 below. However, the test is inconsistent, i.e. the power does not tend to 1 as the sample size passes to infinity. Heuristically, this is due to the fact that, under the alternative, the estimation error of $\hat{\beta}^{bw}$ (rescaled by $\sqrt{n}$) has the extra term

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}f_t'\left(\gamma_i-\gamma\right)\right];$$

under the (quite standard: see e.g. Assumption 3 in Pesaran, 2006) random coefficients assumption for $\gamma_i$, such term has the same order of magnitude as the leading term (thus ruling out power versus local alternatives), and it does not converge to a constant; rather, it can be shown to converge to a normally distributed random variable. This has the effect of inflating the variance of $\sqrt{n}\left(\hat{\beta}^{IE} - \hat{\beta}^{bw}\right)$, but it does not introduce any non-centrality parameter that would diverge under alternatives, whence the result in the theorem.

*Testing for $H_0^b : f_t = f$*

Under $H_0^b$, $\beta$ can be estimated as

$$\hat{\beta}^{wn} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{it}\bar{x}_{it}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{it}\bar{y}_{it}\right),$$

where $\bar{x}_{it} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$ and $\bar{y}_{it} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$; $\hat{\beta}^{wn}$ is the Mean-Group version of the "within" estimator. Based on this, testing for $H_0^b$ could be done using either

$$S_{f,nT}^{IE/CCE} = nT\left(\hat{\beta}^{IE/CCE} - \hat{\beta}^{wn}\right)'\left[Var\left(\hat{\beta}^{IE/CCE} - \hat{\beta}^{wn}\right)\right]^{-1}\left(\hat{\beta}^{IE/CCE} - \hat{\beta}^{wn}\right).$$

It holds that

**Theorem 7** *Let Assumptions 1-4 hold. As $(n,T) \to \infty$, under $H_0^b$*

$$S_{f,nT}^{IE} = O_p(1) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p\left(\sqrt{\frac{n}{T}}\right), \tag{31}$$

$$S_{f,nT}^{CCE} = O_p(1) + O_p\left(\sqrt{\frac{T}{n}}\right). \tag{32}$$

More specifically, as far as $S_{f,nT}^{IE}$ is concerned, equation (31) states that Hausman-type tests based on the IE estimator cannot be employed, as they always diverge under the null. The reason is that, in the expansion of $\tilde{\beta}_i^{IE} - \beta_i$, there are terms of order $O_p\left(n^{-1}\right) + O_p\left(T^{-1}\right)$, which do not get averaged out when calculating the cross-sectional averages. Thus, the impact of such terms on $\sqrt{nT}\left(\hat{\beta}^{IE} - \beta\right)$ is of order $O_p\left(\sqrt{\frac{n}{T}}\right) + O_p\left(\sqrt{\frac{T}{n}}\right)$, which diverges as $(n,T) \to \infty$. As far as $S_{f,nT}^{CCE}$ is concerned, equation (32) states that $S_{f,nT}^{CCE}$ could potentially be employed, at least under the restriction that $\frac{T}{n} \to 0$. As we point out in the proof in Appendix, the problem with this approach is that, in general, the distribution of the $O_p(1)$ term is degenerate, and it anyway depends on several nuisance parameters in the DGP of the $x_{it}$s, and on $f_t$ and $\gamma_i$. In essence, equation (32) states that testing for no factor structure using $S_{f,nT}^{CCE}$ is fraught with difficulties and, in general, not feasible.

# 5 Small sample properties

In this section, we evaluate, through synthetic data, the small sample properties of estimators of $\gamma_i$ and $f_t$ (discussed in Section 2), and the power and size of tests for (13) and (14) based on $S_{\gamma,nT}$ and $S_{f,nT}$ (discussed in Section 3).

The Monte Carlo settings are as follows. Based on model (1)-(2), we consider the following data generating process (DGP):

$$y_{it} = \beta_i x_{it} + \gamma_i f_t + \epsilon_{it}, \tag{33}$$

$$x_{it} = \mu_i + \lambda_i f_t + \epsilon_{it}^x, \tag{34}$$

i.e. we consider model (1)-(2) with $m = r = 1$ - only one individual specific regressor, $x_{it}$, and only one common factor, $f_t$. Unreported simulations show that increasing either $r$ or $m$ does not alter the results. In the simulations, we generate the parameters $\beta_i$ and $\mu_i$ as *i.i.d.* $N(1, 1)$. The common factor $f_t$, the loading $\lambda_i$, and both error terms $\epsilon_{it}$ and $\epsilon_{it}^x$ are all generated as *i.i.d.* $N(0, 1)$ unless otherwise stated. Results are reported for $(n, T) \in \{30, 50, 100, 200\} \times \{30, 50, 100, 200\}$. Finally, in both exercises, simulations are carried out with 5000 iterations.

## 5.1 Small sample properties - $\hat{\gamma}_i$ and $\hat{f}_t$

We evaluate the small sample properties of the estimators $\hat{\gamma}_i$ and $\hat{f}_t$.

As far as $\hat{f}_t$ is concerned, we follow the same logic as in Bai (2003). We compute the correlation coefficient between $\{\hat{f}_t\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$, for each Monte Carlo iteration $j$ - say $\rho_j^f$. We report the average correlation coefficients, i.e. $J^{-1} \sum_{j=1}^J \rho_j^f$, in Table 1 (recall that $J = 5000$).

**[Insert Table 1 somewhere here]**

Table 1 illustrates that the estimated common factor $\hat{f}_t$ is highly correlated with the unobserved common factor $f_t$. This reinforces the results in Bai (2003), albeit obtained in a different context, that the estimated factors are quite good at tracking the true ones; indeed, numerical values are very similar to those in Table 1 in Bai (2003, p.151). When $n$ and $T$ are $\geq 100$, the estimated factors can be treated as the true ones.

As far as $\hat{\gamma}_i$ is concerned, we report confidence intervals for $\gamma_i$. In order to illustrate how confidence intervals shrink as $T$ expands, we set $n = 50$ and $T = 20, 50, 100, 1000$.

According to equation (6) in Theorem 1, as $(n, T) \to \infty$ with $\frac{\sqrt{T}}{n} \to 0$, the 95% confidence interval for $H^{-1}\gamma_i$ is given by $\hat{\gamma}_i \pm \frac{1.96}{\sqrt{T}} \times \hat{\Sigma}_{\gamma i}^{1/2}$. Further, let $\hat{\delta}$ be the least square estimate of $\delta$ in $\Gamma = \hat{\Gamma}\delta + error$, where $\Gamma = (\gamma_1, \ldots \gamma_n)'$ and $\hat{\Gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_n)'$. The 95% confidence interval for $\gamma_i$ is therefore obtained as $\hat{\delta} \times \left( \hat{\gamma}_i \pm \frac{1.96}{\sqrt{T}} \times \hat{\Sigma}_{\gamma i}^{1/2} \right)$. By rotating $\hat{\gamma}_i$ towards $\gamma_i$, we consider the confidence interval for $\gamma_i$ directly, reported in Figure 1.

**[Insert Figure 1 somewhere here]**

Figure 1 shows that, in most cases and for all combinations of $n$ and $T$, the confidence intervals contain the true value of $\gamma_i$. This also holds true for the case $(n, T) = (50, 1000)$, where the ratio $\frac{\sqrt{T}}{n}$

is not negligible, as the theory would require. As predicted by the theory, as $T$ grows, the confidence intervals collapse to the true value of $\gamma_i$.

## 5.2 Small sample properties - $S_{\gamma,nT}$ and $S_{f,nT}$

In this subsection, we report empirical rejection frequencies and power for tests based on the max-type statistics $S_{\gamma,nT}$ and $S_{f,nT}$ defined in (17) and (18) respectively.

As far as the design of the Monte Carlo is concerned, recall that the variance of the common components $c_{it} = \gamma_i f_t$ is set equal to 1 across all experiments. We conduct our simulations for different values of the signal-to-noise ratio $\frac{Var(c_{it})}{\sigma_\epsilon^2}$, where $\sigma_\epsilon^2$ is the variance of $\epsilon_{it}$, equal to $\left\{\frac{1}{3}, \frac{1}{2}, 1\right\}$.

In addition to conducting simulations under the DGP (33), we also consider two alternative DGPs that are nested in (33), in order to assess the robustness of the tests proposed to different specifications of (1)-(2). We firstly consider a DGP for the regressors $x_{it}$ that modifies (34) by not containing common factors, viz.

$$x_{it} = \mu_i + \epsilon_{it}^x. \tag{35}$$

In this case, cross dependence in the $y_{it}$s is purely due to the presence of $f_t$ in (33). The rank condition in Assumption 3*(ii)* does not hold, although the CCE estimator is still consistent. Secondly, we consider a DGP for (1) in which there are no unit specific regressors, viz.

$$y_{it} = \gamma_i f_t + \epsilon_{it}; \tag{36}$$

this is a pure factor model, that fits in the class of models considered by Bai (2003). In this case, it can be argued that testing for no factor structure (either by using $S_{\gamma,nT}$ or $S_{f,nT}$) complements the information criteria in Bai and Ng (2002), by being a test for $r = 0$. This is can also be compared with the framework in Baltagi, Kao, and Na (2012).

Critical values have been computed by approximating $B_n$ and $B_T$ as discussed in Section 3. Unreported simulations show that results worsen only slightly when using the asymptotic critical values.[2]

*Testing for* $H_0^a : \gamma_i = \gamma$

When evaluating the empirical rejection frequencies for tests based on $S_{\gamma,nT}$, we run the Monte Carlo simulations under the null $\gamma_i = 1$ for all $i$. When evaluating power, we generate the loadings

[2] The simulation results are available upon request.

$\gamma_i$ as *i.i.d.* $N\left(1, \sigma_\gamma^2\right)$, reporting results for the case of $\sigma_\gamma = 0.2$. Given that $\epsilon_{it}$ is cross sectionally uncorrelated and homoskedastic by design, $\Sigma_{\gamma i}$ is estimated as $\hat{\Sigma}_{\gamma i} = \hat{\sigma}_\epsilon^2 \times T\left(\hat{F}' M_{x_i} \hat{F}\right)^{-1}$, where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$.

Results for size and power when using the main DGP (33)-(34) are in Table 2.

[**Insert Table 2 somewhere here**]

We firstly consider the empirical rejection frequencies (left panel in the table). The test has a tendency to be oversized in small samples; as a general rule, the correct size is attained when $T \geq 100$ and $n \geq 50$; indeed, when $\sigma_\epsilon^2 = 1$ (high signal-to-noise ratio), the test has satisfactory size properties even for $T = 50$. The Table also shows that, as the signal-to-noise ratio decreases (i.e., as $\sigma_\epsilon^2$ increases), the tendency towards small sample oversizement worsens. This is not so when $T \geq 100$ and $n \geq 50$: the test attains the correct size even for large values of $\sigma_\epsilon^2$.

As far as the power is concerned (right panel in the Table), the test has good power properties in all cases: the power is above 50% for almost all cases. We note that, similarly to the size, the power deteriorates as the signal-to-noise ratio decreases; when $n$ and $T$ are sufficiently large, this disappears.

When considering the two alternative specifications (33)-(35) and (36), results are reported in Tables 3 and 4.

[**Insert Tables 3 and 4 somewhere here**]

Results do not change much with respect to the ones in Table 2, as far as both empirical rejection frequencies and power are concerned. Indeed, the size improves in both cases (especially when simulations are conducted under (36)). When the signal-to-noise ratio is sufficiently high, the test attains its nominal size for all values of $n$, as long as $T \geq 100$.

It is interesting to note that both size and power become much better under (36) than in the other cases. The correct size is attained as long as $n \geq 30$ and $T \geq 50$; moreover, the power is always above 90% for all combinations of $n$ and $T$.

*Testing for* $H_0^b : f_t = f$

We run the Monte Carlo simulations under the null $f_t = 1$ for all $t$ when evaluating the size of tests based on $S_{f,nT}$. When evaluating the power, we generate the common factors $f_t$ as *i.i.d.* $N\left(1, \sigma_f^2\right)$, reporting results for the case of $\sigma_f = 0.2$. Finally, we estimate $\Sigma_{ft}$ as $\Sigma_{ft} = V_{nT}^{-1}\hat{\sigma}_\epsilon^2 \frac{1}{n}\sum_{i=1}^n \hat{\lambda}_i \hat{\lambda}_i' V_{nT}^{-1}$ where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$.

Results when using (33)-(34) are in Table 5.

**[Insert Table 5 somewhere here]**

The size of the test is almost always the correct one, with few exceptions - the test is oversized for small $T$ when $\sigma_\epsilon^2$ is high. Both $n$ and $T$ have a quite limited impact on the results.

The test has very good power properties, especially when the signal-to-noise ratio is high. We note that the power increases with both $n$ and $T$, in a more pronounced way with $n$.

As in the previous subsection, we also considered size and power under the alternative specifications (33)-(35) and (36); results are in Tables 6 and 7.

**[Insert Tables 6 and 7 somewhere here]**

Results do not differ much, when carrying out simulations under (33)-(35), from the values in Table 5. Actually, as it was noted for the case of $S_{\gamma,nT}$, results improve slightly, in particular the power. Similar considerations hold for the empirical rejection frequencies computed under (36): the size is always the correct one. The power is also very good, under all possible combinations of parameters.

For the sake of completeness, we run both tests using as a first step estimator the IE proposed by Song (2013). The size and power reported in Table 8, for the $S_\gamma$ test, when the DGP is the one in equations (33)-(34), show that the test procedure is unaffected by the choice of the first step estimator when this is a consistent one.

**[Insert Table 8 somewhere here]**

*Autocorrelated and heteroskedastic idiosyncratic errors*

In order to assess the finite sample properties of the two test procedures when the errors are autocorrelated and heteroskedastic, we consider the following DGP:

$$\epsilon_{it} = 0.5\epsilon_{it-1} + u_{it}$$

$$u_{it} \sim IIDN(0, \sigma_{ui}^2) \quad \sigma_{ui}^2 \sim U(0.1, 0.5)$$

and we make use of the HAC estimators for $\Sigma_\gamma$ and $\Sigma_f$ given by equations (7) , (10). Apart from these features, the experiments have the same specifications as above. As far as the noise-to-signal ratio is concerned, results are very similar to the *i.i.d.* cases, and we only report the cases in which $\sigma_\epsilon^2 = 1$ (i.e. the worst case, based on the simulations above) to save space.

**[Insert Tables 9 and 10 somewhere here]**

The results in Tables 9 and 10 can be compared with the *i.i.d.* cases in Tables 2 and 5 respectively. In the case of non *i.i.d.* errors, both tests have a tendency to be oversized in small samples, $(n, T) \leq 50$. However, as both dimensions are larger than 50, the empirical rejection frequencies become almost undistinguishable from the ones computed with *i.i.d.* errors. As far as, the power is concerned, both tests have good properties and are very close to the *i.i.d.* case.

# 6 Conclusions

In this contribution, we develop an inferential theory for the unobservable common factors and their loadings in a large, stationary panel model with observable regressors. Our framework allows for slope heterogeneity; we also allow for correlation between common factors and observable regressors, by modelling the DGP of the observable regressors as containing the common factors, in a similar spirit as in Pesaran (2006).

We extend the framework in Pesaran (2006) by providing a two stage estimator for the unobserved common factors and their loading. We derive rates of convergence and limiting distribution of both the estimated factors and loadings, using a similar method of proof to Bai (2009a). In a similar vein to Sarafidis, Yamagata and Robertson (2009), we also develop two tests for the null of no factor structure, based on the null that factor loadings are homogeneous, and that common factors are homogeneous over time, respectively. In either case, the assumed factor model boils down to a model with (time specific or unit specific) common effects, so that common features in the panel can be captured by inserting time dummies or unit specific dummies. The proposed test procedures simplify the specification analysis of heterogeneous panel data models with unobserved factors. From a methodological perspective, this entails that the tests can be implemented without prior knowledge of the number of factors. The only thing which is needed is a consistent preliminary estimation of the slope parameters. Building on this, we propose statistics based on extrema of the estimated loadings and common factors. Under the null, the test statistics converge to an Extreme Value distribution. As far as power is concerned, from a theoretical point of view our tests are consistent even under alternatives where only one loading or common factor differs from the average. Monte Carlo evidence shows that both tests have the correct size and good power properties.

Building on the theory developed in this paper, there are several interesting avenues for further developments. An important case is the estimator of the $\beta_i$s used in Step 1. In our paper, we focus on the CCE estimator proposed by Pesaran (2006); this estimator is easy to treat analytically, but it is only a possible choice. In particular, our setup requires strict exogeneity, thereby ruling out e.g. the possibility of having lagged values of the $y_{it}$s among the regressors. This requirement is due to the estimation method employed in Step 1, rather than to the inference on factors and loadings per se. Indeed, the CCE is known not to work in presence of weakly exogenous regressors (see Everaert and Groote, 2012; and Chudik and Pesaran, 2013). However, the assumption of strict exogeneity can be readily relaxed (accommodating e.g. for dynamic models), upon employing, in Step 1, an estimator of the $\beta_i$s that is consistent at a rate $O_p\left[\min\left\{T^{-1/2}, n^{-1}\right\}\right]$. A possible choice for this case is the IE estimator studied in Song (2013), which has the desired convergence rate, even in presence of dynamic models. Alternatively, a different approach, based on unit specific estimators can be used, by instrumenting the unobservable common factors $f_t$ using the regressors $x_{jt}$ for each unit $i$, with $i \neq j$ - indeed, both the CCE and the IE have a natural Instrumental Variable interpretation (see also Bai, 2009b). Such extensions are currently under investigation of the authors.

**Acknowledgement**

This is a revised version of a paper previously circulated under the working title "Two-Stage Inference in Heterogeneous Panels". We are very grateful to the Editor (Cheng Hsiao), one anonymous Associate Editor and two anonymous Referees for very constructive feedback which has greatly improved the generality of the paper. We also wish to thank the participants to the Faculty of Finance Workshops at Cass Business School; to the New York Camp Econometrics V (Syracuse University, NY, October 2010); to the 4th CSDA International Conference on Computational and

Financial Econometrics (London, December 2010); to the 18th International Conference on Panel Data (Banque de France, July 2012), in particular Chihwa Kao, Jean-Pierre Urbain and Takashi Yamagata. Special thanks go to Lajos Horvath for providing us with valuable comments. The usual disclaimer applies.

# Appendix A: Technical Lemmas

In this Appendix and the next one, we set $H = I_r$ in the proofs (although not in the statements of the Lemmas), for the sake of notational simplicity. Inequalities are written, when possible, omitting constants.

The Lemmas in this Section extend various results in Bai (2009a,b) to our framework. All proofs rely upon the decomposition - see Proposition A.1 in Bai (2009a):

$$\begin{aligned}\hat{F}-F &= \frac{1}{nT}\sum_{j=1}^{n}X_j\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'X_j'\hat{F} \\ &-\frac{1}{nT}\sum_{j=1}^{n}X_j\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'F'\hat{F}-\frac{1}{nT}\sum_{j=1}^{n}X_j\left(\tilde{\beta}_j-\beta_j\right)\epsilon_j'\hat{F} \\ &-\frac{1}{nT}\sum_{j=1}^{n}F\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'X_j'\hat{F}-\frac{1}{nT}\sum_{j=1}^{n}\epsilon_j\left(\tilde{\beta}_j-\beta_j\right)'X_j'\hat{F} \\ &+\frac{1}{nT}\sum_{j=1}^{n}F\gamma_j\epsilon_j'\hat{F}+\frac{1}{nT}\sum_{j=1}^{n}\epsilon_j\gamma_j'F'\hat{F}+\frac{1}{nT}\sum_{j=1}^{n}\epsilon_j\epsilon_j'\hat{F}.\end{aligned} \tag{37}$$

In (37), the main difference with Bai (2009a) is the presence of the unit specific estimates, $\tilde{\beta}_j$. Consider also the following notation, which we use henceforth throughout Appendices A and B. We define $\Upsilon_i \equiv \left(X_i'\bar{M}_wX_i\right)^{-1}\left(X_i'\bar{M}_w\epsilon_i\right)$, so that we can write

$$\begin{aligned}\tilde{\beta}_i-\beta_i &= \left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_w\epsilon_i}{T}\right)+\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_wF}{T}\gamma_i\right) \\ &= \Upsilon_i+\bar{\Upsilon}_i,\end{aligned} \tag{38}$$

for every $i$; by construction, $\bar{\Upsilon}_i = O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)$. We extensively use the notation $\delta_{nT} = \min\left\{\sqrt{n},\sqrt{T}\right\}$ and $\phi_{nT} = \min\left\{n,\sqrt{T}\right\}$.

**Lemma A.1** *Under Assumptions 1-4, it holds that, for every $i$, $E\left\|\tilde{\beta}_i-\beta_i\right\|^r = O\left(\phi_{nT}^{-r}\right)$, for any $r \leq 3$.*

**Proof.** Let $\|A\|_1$ denote the $L_1$-norm of a matrix $A$, i.e. $\|A\|_1 = \max_{x\neq 0}\|Ax\|_1/\|x\|_1$. By a well known norm inequality (see e.g. Strang, 1988, p. 369, exercise 7.2.3), it holds that

$$\begin{aligned}\left\|\tilde{\beta}_i-\beta_i\right\|^r &\leq \left\|\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\right\|_1^r\left\|\frac{X_i'\bar{M}_w\epsilon_i}{T}+\frac{X_i'\bar{M}_wF}{T}\gamma_i\right\|^r \\ &= \left[l_{\min}^{-1}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)\right]^r\left\|\frac{X_i'\bar{M}_w\epsilon_i}{T}+\frac{X_i'\bar{M}_wF}{T}\gamma_i\right\|^r,\end{aligned}$$

where the last equality holds by symmetry. In view of Assumption 4*(i)*, and omitting $\gamma_i$ by virtue of Assumption 3*(iii)*

$$E\left\|\tilde{\beta}_i-\beta_i\right\|^r \leq E\left\|\frac{X_i'\bar{M}_w\epsilon_i}{T}\right\|^r + E\left\|\frac{X_i'\bar{M}_wF}{T}\right\|^r = I + II.$$

Consider $I$; we have $I \leq T^{-r}E\left\|X_i'\epsilon_i\right\|^r = T^{-r}E\left\|\sum_{t=1}^T x_{it}\epsilon_{it}\right\|^r$. It holds that

$$\begin{aligned}
T^{-r}E\left\|X_i'\epsilon_i\right\|^r &\leq T^{-r}E\left|\sum_{t=1}^T\left\|x_{it}\epsilon_{it}\right\|^2\right|^{r/2} \leq T^{-r}E\left|T^{1-2/r}\left(\sum_{t=1}^T\left\|x_{it}\epsilon_{it}\right\|^r\right)^{2/r}\right|^{r/2} \qquad (39)\\
&\leq T^{-r}T^{r/2}\left(\frac{1}{T}\sum_{t=1}^T E\left\|x_{it}\epsilon_{it}\right\|^r\right) \leq T^{-r/2}\left(\frac{1}{T}\sum_{t=1}^T\left[E\left\|x_{it}\right\|^{2r}\right]^{1/2}\left[E\left|\epsilon_{it}\right|^{2r}\right]^{1/2}\right)\\
&= O\left(T^{-r/2}\right),
\end{aligned}$$

where we have used: Assumption 2*(iv)*; Holder's inequality; the $C_r$-inequality and Jensen's inequality; the Cauchy-Schwartz inequality; and the fact that, by Assumptions 1 and 2*(i)*, $E\left|\epsilon_{it}\right|^{2r} < \infty$ and $E\left\|x_{it}\right\|^{2r} < \infty$ respectively. Using the Cauchy-Schwartz inequality in this context is more than what is necessary, since $x_{it}$ and $\epsilon_{it}$ are independent. Turning to $II$, note that, for sufficiently large $n$ and omitting higher order terms, $\left(\bar{H}_w'\bar{H}_w\right)^{-1} = D_w^{-1} - D_w^{-1}R_wD_w^{-1}$, with $D_w = C'F'FC$ and $\left\|R_w\right\| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)$ - see e.g. equation (29) in Pesaran (2006). Therefore, letting $\bar{\epsilon} = n^{-1}\sum_{i=1}^n\epsilon_i$ and omitting higher order terms

$$\begin{aligned}
\frac{X_i'\bar{M}_wF}{T} &= -\frac{X_i'FCD_w^{-1}\bar{\epsilon}'F}{T^2} - \frac{F'\bar{\epsilon}D_w^{-1}C'F'X_i}{T^2} \qquad (40)\\
&\quad -\frac{X_i'\bar{\epsilon}D_w^{-1}\bar{\epsilon}'F}{T^2} - \frac{X_i'FCD_w^{-1}R_wD_w^{-1}C'F'F}{T^2}\\
&= -I - I' - II - III.
\end{aligned}$$

Consider $E\left\|I\right\|^r$; since $C$ has full rank by Assumption 4*(ii)* and $D_w$ is invertible

$$E\left\|I\right\|^r \leq E\left\|\frac{X_i'F}{T}\frac{\bar{\epsilon}'F}{T}\right\|^r \leq M\left[E\left\|\frac{X_i'F}{T}\right\|^{2r}\right]^{1/2}\left[E\left\|\frac{\bar{\epsilon}'F}{T}\right\|^{2r}\right]^{1/2}.$$

Consider the first term; we have $T^{-1}\sum_{t=1}^T E\left\|x_{it}f_t'\right\|^{2r} \leq T^{-1}\sum_{t=1}^T\left[E\left\|x_{it}\right\|^{4r}\right]^{1/2}\left[E\left\|f_t\right\|^{4r}\right]^{1/2}$, which is finite by Assumption 2*(i)*. As far as the second term is concerned, note

$$E\left\|\frac{\bar{\epsilon}'F}{T}\right\|^{2r}\leq T^{-2r}\sum_{t=1}^{T}\left[E\left\|f_t\right\|^{4r}\right]^{1/2}\left[E\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_{it}\right|^{4r}\right]^{1/2},$$

after similar passages as in equation (39). It holds that $E\left\|f_t\right\|^{4r}<\infty$ by Assumption 2*(i)*. By using Assumption 1*(iv)*(b) and following thereafter a similar logic as in the proof of (39), we have $E\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_{it}\right|^{4r}=O\left(n^{-r/2}\right)$, so that $E\left\|I\right\|^{r}=O\left(n^{-r/2}T^{-r/2}\right)$. The same logic yields $E\left\|II\right\|^{r}=O\left(n^{-r}T^{-r}\right)$. Finally, consider $III$; after some passages

$$E\left\|III\right\|^{r}\leq\left\|R_w\right\|^{r}\left[E\left\|\frac{X_i'F}{T}\right\|^{2r}\right]^{1/2}\left[E\left\|\frac{F'F}{T}\right\|^{2r}\right]^{1/2}=O\left(\left\|R_w\right\|^{r}\right),$$

again by similar passages as above. Therefore, $E\left\|\frac{X_i'\bar{M}_wF}{T}\right\|^{r}=O\left(\left\|R_w\right\|^{r}\right)$. Putting everything together, the Lemma follows. QED

**Lemma A.2** *Under Assumptions 1-4, it holds that, for every $i$*

**A.2*(i)*** $T^{-1}\epsilon_i'\left(\hat{F}-F\right)=O_p\left(\delta_{nT}^{-2}\right)$;

**A.2*(ii)*** $n^{-1/2}T^{-1}\sum_{i=1}^{n}\epsilon_i'\left(\hat{F}-F\right)=O_p\left(n^{-1/2}\right)+O_p\left(T^{-1}\right)$.

**Proof.** The proof of A.2*(i)* is very similar, and in fact simpler, than that of A.2*(ii)*; thus we focus on the latter only. Using (37)

$$\begin{aligned}
& n^{-1/2}T^{-1}\sum_{i=1}^{n}\epsilon_i'\left(\hat{F}-F\right) \qquad (41)\\
= & \frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'X_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}-\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'X_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'\hat{F}}{T}\\
& -\frac{1}{nT}\sum_{j=1}^{n}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'X_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)\frac{\epsilon_j'\hat{F}}{\sqrt{T}}-\frac{1}{n\sqrt{T}}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'F}{\sqrt{T}}\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\\
& -\frac{1}{\sqrt{n}}\frac{1}{nT}\sum_{j=1}^{n}\sum_{i=1}^{n}\epsilon_i'\epsilon_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}+\frac{1}{\sqrt{nT}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\epsilon_i'F}{\sqrt{T}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma_j\frac{\epsilon_j'\hat{F}}{T}\\
& +\frac{1}{\sqrt{nT}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i'\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon_j\gamma_j'\frac{F'\hat{F}}{T}+\frac{1}{T}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'\epsilon_j\frac{\epsilon_j'\hat{F}}{T}\\
= & I+II+III+IV+V+VI+VII+VIII.
\end{aligned}$$

The proof follows very similar lines to that of Lemma A.8 in Song (2013): the only difference is the different expansion of the estimation error $\tilde{\beta}_j - \beta_j$ when using the CCE. Thus, we report only the complete passages to determine the order of magnitude of $I$; the same logic applies to all the other terms in the expansion. The only term for which passages slightly differ is $V$, and we report the full blown proof for it.

Consider $I$; it holds that $I \leq n^{-1} \sum_{j=1}^{n} \left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\| \left\| \frac{X_j' \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2$. This is bounded by

$$
\begin{aligned}
& E\left[ \left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\| \left\| \frac{X_j' \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2 \right] \qquad (42) \\
\leq \; & \left[ E\left( \left\| \tilde{\beta}_j - \beta_j \right\|^{2p} \right) \right]^{1/p} \left[ E\left( \left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\| \left\| \frac{X_j' \hat{F}}{T} \right\| \right)^q \right]^{1/q} \\
\leq \; & \left[ E\left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{2/3} \left[ E\left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\|^6 \right]^{1/6} \left[ E\left\| \frac{X_j' \hat{F}}{T} \right\|^6 \right]^{1/6},
\end{aligned}
$$

using Holder's inequality in the first line (with $p = \frac{3}{2}$ and $q = 3$), and the Cauchy-Schwartz inequality in the second line. The first term is of order $O\left(\phi_{nT}^{-2}\right)$ in light of Lemma A.1. Similar passages as in the proof of Lemma A.1 yield that both the second and third terms are of order $O(1)$. This entails that $I = O_p\left(T^{-1/2}\phi_{nT}^{-2}\right)$. Similar passages yield $II = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$; $III = O_p\left(T^{-1}\phi_{nT}^{-1}\right)$; $IV = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$; $VI = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right) + O_p\left(\delta_{nT}^{-2}\right)$; $VII = O_p\left(n^{-1/2}\right)$ and $VIII = O_p\left(T^{-1}\right)$ $+O_p\left(n^{-1/2}T^{-1/2}\right)$.

Consider now $V$, whose proof is marginally different to that of Song (2013)

$$
\begin{aligned}
V \; \leq \; & \left[ \frac{1}{n} \sum_{j=1}^{n} E\left| \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \right]^{1/2} \left[ \frac{1}{n} \sum_{j=1}^{n} E\left( \left\| \frac{X_j' \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\| \right)^2 \right]^{1/2} \\
\leq \; & \left[ \frac{1}{n} \sum_{j=1}^{n} E\left| \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \right]^{1/2} \left[ \frac{1}{n} \sum_{j=1}^{n} E\left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{1/3} \left[ \frac{1}{n} \sum_{j=1}^{n} E\left\| \frac{X_j' \hat{F}}{T} \right\|^6 \right]^{1/6} \\
= \; & \left[ \frac{1}{nT} \sum_{j=1}^{n} E\left| \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \right]^{1/2} O_p\left(\phi_{nT}^{-1}\right),
\end{aligned}
$$

using the Cauchy-Schwartz inequality (first line), Holder's inequality with the same orders as in (42) (second line) and Lemma A.1. Also, $E\left| \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \leq (nT)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T}$ $\left| E\left(\epsilon_{it}\epsilon_{kt}\epsilon_{js}\epsilon_{ks}\right) \right| \leq M$, by Assumption 1*(iii)*(d), so that $V = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Putting all together, part A.2*(ii)* follows. QED

**Lemma A.3.** *It holds that, for every $i$*

**A.3***(i)* $T^{-1}X_i'\left(\hat{F}-FH\right)=O_p\left(\delta_{nT}^{-2}\right)$;

**A.3***(ii)* $T^{-1}F'\left(\hat{F}-FH\right)=O_p\left(\delta_{nT}^{-2}\right)$;

**A.3***(iii)* $T^{-1}\left(\hat{F}-FH\right)'\left(\hat{F}-FH\right)=O_p\left(\delta_{nT}^{-2}\right)$.

**Proof**. The Lemma is a refinement of Lemma A.3 in Bai (2009a). Particularly, by Lemma A.3 in Bai (2009a) we have $T^{-1}X_i'\left(\hat{F}-F\right)=o_p(1)$ and $T^{-1}F'\left(\hat{F}-F\right)=o_p(1)$.

Consider part *(i)*. Using (37)

$$
\begin{aligned}
\frac{X_i'\left(\hat{F}-F\right)}{T} &= \frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T} \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'\hat{F}}{T}-\frac{1}{nT}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\epsilon_j'\hat{F} \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'F}{T}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}-\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T} \\
&\quad +\frac{1}{nT}\sum_{j=1}^{n}\frac{X_i'F}{T}\gamma_j\epsilon_j'\hat{F}+\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\gamma_j'\frac{F'\hat{F}}{T}+\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\frac{\epsilon_j'\hat{F}}{T} \\
&= I-II-III-IV-V+VI+VII+VIII;
\end{aligned}
$$

henceforth, we omit $\gamma_i$ in the passages, based on Assumption 3*(iii)*. Consider $I$; it is bounded by $E\left(\left\|\frac{X_i'X_j}{T}\right\|\left\|\frac{X_j'\hat{F}}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|^2\right)$. Using the Holder's inequality and the Cauchy-Schwartz inequality in a similar way to (42), this is bounded by $\left[E\left\|\frac{X_i'X_j}{T}\right\|^6\right]^{1/6}\left[E\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/6}\left[E\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{2/3}=O_p\left(\phi_{nT}^{-2}\right)$. Turning to $II$, we have $II=\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'F}{T}+o_p(1)$, where the $o_p(1)$ term comes from $T^{-1}F'\left(\hat{F}-F\right)=o_p(1)$. By (38), this is bounded by $\left\|\frac{X_i}{\sqrt{T}}\right\|\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\Upsilon_j\right\|+\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{X_i'X_j}{T}\right\|\left\|\bar{\Upsilon}_j\right\|=II_a+II_b$. Consider $II_a$; since

$$
\begin{aligned}
E\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\Upsilon_j\right\|^2 &= E\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\right\|^2 \\
&= \frac{1}{n^2T}\sum_{j=1}^{n}E\left\|X_j\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\right\|^2 \\
&\leq \frac{1}{n^2T}\sum_{j=1}^{n}\left[E\left\|\frac{X_j}{\sqrt{T}}\right\|^2\right]^{1/2}\left[E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2\right]^{1/2}=O\left(\frac{1}{nT}\right),
\end{aligned}
$$

where we have used Assumptions 4*(i)*, 1, 2*(i)*; the Cauchy-Schwartz inequality; and the facts that both $E\left\|\frac{X_j}{\sqrt{T}}\right\|^2$ and $E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2$ are finite. The latter statement can be shown as follows

$$E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2 = E\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{jt}x_{js}'\epsilon_{jt}\epsilon_{js}\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|x_{jt}x_{js}'\right\|E\left|\epsilon_{jt}\epsilon_{js}\right| \leq M\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{jt}\epsilon_{js}\right| \leq M',$$

using, respectively, Assumptions 2*(iii)*, 2*(i)* and 1*(ii)*(c). Thus, $II_a = O_p\left(n^{-1/2}T^{-1/2}\right)$. Turning to $II_b$, this is of the same order of magnitude as $E\left[\left\|\frac{X_i'X_j}{T}\right\|\left\|\bar{\Upsilon}_j\right\|\right] \leq \left[E\left(\left\|\frac{X_i'X_j}{T}\right\|^2\right)\right]^{1/2}\left[E\left(\left\|\bar{\Upsilon}_j\right\|^2\right)\right]^{1/2}$; by the proof of Lemma A.1, $II_b = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Thus, $II = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Turning to $III$, it can be decomposed into $\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\frac{\epsilon_j'F}{\sqrt{T}} + \frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\frac{\epsilon_j'\left(\hat{F}-F\right)}{T} = III_a + III_b$; $III_a$ is bounded by $\frac{1}{\sqrt{T}}E\left[\left\|\frac{X_i'X_j}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|\right]$. Using the same logic as above, this entails $III_a = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Similar passages and Lemma A.2*(i)*, yield $III_b = O_p\left(\phi_{nT}^{-1}\delta_{nT}^{-2}\right)$. Term $IV$ has the same magnitude of term $II$. As far as $V$ is concerned, using the fact that $T^{-1}X_i'\left(\hat{F}-F\right) = o_p(1)$, $V$ is bounded by $T^{-1/2}E\left[\left\|\frac{X_i'\epsilon_j}{\sqrt{T}}\right\|\left\|\frac{X_j'F}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|\right]$; a similar logic to the proof of $I$ yields $V = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Turning to $VI$, we have $VI = \frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'F}{T}\frac{\epsilon_j'F}{\sqrt{T}} + \frac{X_i'F}{T}\frac{1}{n}\sum_{j=1}^{n}\frac{\epsilon_j'\left(\hat{F}-F\right)}{T} = VI_a + VI_b$. Considering $VI_a$, similar passages as above give $VI_a = O_p\left(n^{-1/2}T^{-1/2}\right)$. Turning to $VI_b$, this is $O_p\left(\delta_{nT}^{-2}\right)$ by Lemma A.2*(ii)*. Therefore, $VI = O_p\left(\delta_{nT}^{-2}\right)$. As far as $VII$ is concerned, it is bounded by $\left\|\frac{F'F}{T}\right\|\left\|\frac{1}{nT}\sum_{j=1}^{n}X_i'\epsilon_j\right\| + o_p(1)$. The term $\sum_{j=1}^{n}X_i'\epsilon_j$ is bounded by the square root of its variance, viz.

$$\begin{aligned} E\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}x_{is}'\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\epsilon_{jt}\epsilon_{ks}\right)\right\| &\leq \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|x_{it}x_{is}'\right\|E\left|\epsilon_{jt}\epsilon_{ks}\right| \\ &\leq M\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{jt}\epsilon_{ks}\right| \leq M'(nT). \end{aligned}$$

Thus, $VII = O_p\left(n^{-1/2}T^{-1/2}\right)$. Finally, $VIII = \frac{1}{nT}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\frac{\epsilon_j'F}{\sqrt{T}} + \frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\frac{\epsilon_j'\left(\hat{F}-F\right)}{T} = VIII_a + VIII_b$; $VIII_a$ is bounded by $\left[E\left\|\frac{X_i'\epsilon_j}{\sqrt{T}}\right\|^2\right]^{1/2}\left[E\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|^2\right]^{1/2}$, which is $O(1)$, so that $VIII_a = O_p\left(T^{-1}\right)$. Similarly, $VIII_b$ is bounded by $T^{-1/2}\left[E\left(\left\|\frac{X_i'\epsilon_j}{\sqrt{T}}\right\|^2\right)\right]^{1/2}\left[E\left(\left\|\frac{\epsilon_j'\left(\hat{F}-F\right)}{T}\right\|^2\right)\right]^{1/2} = O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$, using Lemma A.2*(i)*. Putting all together, part *(i)* of the Lemma follows. The proof of part *(ii)* follows essentially the same passages, and is therefore omitted. As far as part *(iii)* is concerned, the same logic as above can be applied directly to (37), obtaining $T^{-1/2}\left\|\hat{F}-F\right\| = O_p\left(\left\|\tilde{\beta}_j-\beta_j\right\|\right) + O_p\left(\delta_{nT}^{-1}\right)$, whence $T^{-1}\left\|\hat{F}-F\right\|^2 = O_p\left(\delta_{nT}^{-2}\right)$. QED

**Lemma A.4** *Let Assumptions 1-4 hold. Under $H_0^a$ that $\gamma_i = \gamma$, it holds that $\widehat{\overline{\gamma}} - \gamma = O_p\left(\delta_{nT}^{-2}\right)$ as $(n, T) \to \infty$.*

**Proof**. By definition, under $H_0^a$

$$\sqrt{T}\left(\hat{\gamma}_i - \gamma\right) = \left(\frac{\hat{F}' M_{Xi} \hat{F}}{T}\right)^{-1} \left[\frac{\hat{F}' M_{Xi} \epsilon_i}{\sqrt{T}} - \frac{\hat{F}' M_{Xi}\left(\hat{F} - F\right)\gamma}{\sqrt{T}}\right]; \tag{43}$$

also, under $H_0^a$, it holds that $\widehat{\overline{\gamma}} - \gamma = \frac{1}{n}\sum_{i=1}^n \left(\hat{\gamma}_i - \gamma\right)$. Using (43) and neglecting higher order terms coming from $\hat{F}' M_{Xi} \hat{F} - F' M_{Xi} F$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \left(\hat{\gamma}_i - \gamma\right) &= \frac{1}{n}\sum_{i=1}^n \left(F' M_{Xi} F\right)^{-1} F' M_{Xi} \epsilon_i + \frac{1}{n}\sum_{i=1}^n \left(F' M_{Xi} F\right)^{-1} \left(\hat{F} - F\right)' M_{Xi} \epsilon_i \\
&\quad + \frac{1}{n}\sum_{i=1}^n \left(F' M_{Xi} F\right)^{-1} \hat{F}' M_{Xi} \left(\hat{F} - F\right)\gamma \\
&= I + II + III.
\end{aligned}$$

Term $I$ is bounded by the square root of

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n E\left\|\left(\frac{F' M_{Xi} F}{T}\right)^{-1}\left(\frac{F' M_{Xi} \epsilon_i}{T}\right)\left(\frac{F' M_{Xj} F}{T}\right)^{-1}\left(\frac{F' M_{Xj} \epsilon_j}{T}\right)\right\| \\
\leq\ & M \frac{1}{n^2 T^2}\sum_{i=1}^n \sum_{j=1}^n E\left\|\left(F' \epsilon_i\right)\left(F' \epsilon_j\right)\right\| \leq M \frac{1}{n^2 T^2}\sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \sum_{s=1}^T E\left\|f_t f_s'\right\| E\left|\epsilon_{it}\epsilon_{js}\right| \\
\leq\ & M' \frac{1}{n^2 T^2}\sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \sum_{s=1}^T E\left|\epsilon_{it}\epsilon_{js}\right| \leq M'' \frac{1}{nT},
\end{aligned}$$

using, respectively, Assumptions 4*(i)*, 2*(iii)*, 2*(i)* and 1*(ii)*(c). Thus, $I = O_p\left(\frac{1}{\sqrt{nT}}\right)$. Consider $II$; it is bounded by

$$E\left\|\left(\frac{F' M_{Xi} F}{T}\right)^{-1} \frac{\left(\hat{F} - F\right)' M_{Xi} \epsilon_i}{T}\right\| \leq E\left\|\frac{\epsilon_i'\left(\hat{F} - F\right)}{T}\right\| = O_p\left(\delta_{nT}^{-2}\right),$$

by Assumption 3*(i)* and Lemma A.2*(i)*. Similarly, $III$ is bounded by $E\left\|\hat{F}'\left(\hat{F} - F\right)\right\| = O_p\left(\delta_{nT}^{-2}\right)$, by Lemma A.3*(ii)*. The bounds for $II$ and $III$ are not necessarily the sharpest ones, but are sufficient for our purpose. Putting all together, $\widehat{\overline{\gamma}} = \gamma + O_p\left(\delta_{nT}^{-2}\right)$. QED

**Lemma A.5** *Let Assumptions 1-4 hold. Under $H_0^b$ that $f_t = f$, it holds that $\widehat{\overline{f}} - f = O_p\left(\delta_{nT}^{-2}\right)$ as $(n, T) \to \infty$.*

**Proof**. Consider (37) and let $\bar{F} = f \times i_T$; under $H_0^b$, it holds that

$$
\begin{aligned}
\widehat{\hat{F}} - \bar{F} &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j - \beta_j\right)\left(\tilde{\beta}_j - \beta_j\right)'\frac{1}{T}\sum_{t=1}^{T}x_{jt} \\
&\quad -\frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j - \beta_j\right)'\frac{1}{T}\sum_{t=1}^{T}x_{jt} - \frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j - \beta_j\right)'\frac{1}{T}\sum_{t=1}^{T}x_{jt} \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j - \beta_j\right)\gamma_j'\frac{1}{T}\sum_{t=1}^{T}f_t - \frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j - \beta_j\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt} \\
&\quad +\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\gamma_j'\frac{1}{T}\sum_{t=1}^{T}f_t + \frac{1}{n\sqrt{T}}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt} + \frac{1}{nT^{3/2}}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt} \\
&= I - II - III - IV - V + VI + VII + VIII.
\end{aligned}
$$

Consider $I$; it is bounded by $E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|\left\|\tilde{\beta}_j - \beta_j\right\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}x_{jt}\right\|\right] \leq E\left[\left\|\tilde{\beta}_j - \beta_j\right\|^3\right]^{2/3} E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|^6\right]^{1/6}$ $E\left[\left\|\frac{1}{T}\sum_{t=1}^{T}x_{jt}\right\|^6\right]^{1/6} = O\left(\phi_{nT}^{-2}\right)$, using a similar logic to (42) and Lemma A.1. Similar arguments yield $II = O_p\left(n^{-1/2}T^{-1/2}\right) + O_p\left(n^{-1}\right)$, $III = O_p\left(\phi_{nT}^{-2}\right)$, $IV = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$, and $VI = O_p\left(n^{-1/2}T^{-1/2}\right)$. Consider $V$; this is bounded by

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|\left\|\tilde{\beta}_j - \beta_j\right\|\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt}\right|\right] \\
\leq\ & \frac{1}{\sqrt{T}}\left[E\left\|\frac{\hat{F}'X_j}{T}\right\|^4\right]^{1/4}\left[E\left\|\tilde{\beta}_j - \beta_j\right\|^2\right]^{1/2}\left[E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt}\right|^4\right]^{1/4} = \frac{1}{\sqrt{T}}O(1)\,O\left(\phi_{nT}^{-1}\right)O(1),
\end{aligned}
$$

using again Lemma A.1 and the fact that $E\left\|T^{-1/2}\sum_{t=1}^{T}\epsilon_{jt}\right\|^4 = O(1)$ - this can be shown using Assumption 1*(iv)*(b) and similar passages as in the proof of Lemma A.1. Hence, $V = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$; similarly, $VII = O_p\left(n^{-1/2}T^{-1/2}\right)$ and $VIII = O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$. Putting all together, this yields $\widehat{\hat{F}} - \bar{F} = O_p\left(\delta_{nT}^{-2}\right)$. QED

**Lemma A.6** *Let Assumptions 1-4 hold, and let $k$ denote the largest finite moment of $\epsilon_{it}$, $f_t$ and $x_{it}$. It holds that*

**A.6*(i)*** $\max_{1\leq i\leq n}\left\|\tilde{\beta}_i - \beta_i\right\|^2 = o_p\left(n^{2/k}\phi_{nT}^{-2}\right)$;

**A.6*(ii)*** $\max_{1\leq t\leq T}\left\|\hat{f}_t - H'f_t\right\|^2 = o_p\left(T^{2/k}\delta_{nT}^{-2}\right)$;

**A.6*(iii)*** $\max_{1\leq i\leq n}\left\|\hat{\gamma}_i - H^{-1}\gamma_i\right\|^2 = o_p\left(n^{2/k}T^{-1}\right) + o_p\left(n^{2/k-2}T\right)$;

**A.6*(iv)*** $\max_{1\leq t\leq T}\hat{\epsilon}_{it}^2 = o_p\left(T^{2/k}\right) + o_p\left(T^{2/k}\delta_{nT}^{-2}\right)$;

**A.6** *(v)* $\max_{1\leq t\leq T}\left|\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right|=o_p\left(T^{2/k}\delta_{nT}^{-2}\right)$;

**A.6** *(vi)* $\max_{1\leq i\leq n}\hat{\epsilon}_{it}^2=o_p\left(n^{2/k}\right)+o_p\left(n^{4/k}\phi_{nT}^{-2}\right)+o_p\left(n^{2/k-2}T\right)$;

**A.6** *(vii)* $\max_{1\leq i\leq n}\left|\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right|=o_p\left(n^{4/k}\phi_{nT}^{-2}\right)+o_p\left(n^{2/k-2}T\right)$.

**Proof.** In the proof, we extensively use the fact that, for an arbitrary sequence of random variables $Z_1,...,Z_m$ such that $\max_{1\leq h\leq m}E\left|Z_h\right|^a\leq M$ for some $a>0$, it holds that

$$\max_{1\leq h\leq m}\left|Z_h\right|=o_p\left(m^{1/a}\right). \tag{44}$$

The proofs are rather repetitive, and where possible we only provide an intuition of the main argument, omitting passages.

Consider part *(i)*. We know, from the proof of Lemma A.1, that

$$\left\|\tilde{\beta}_i-\beta_i\right\|^2\leq\left\|\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\right\|_1^2\left\|\frac{X_i'\bar{M}_w\epsilon_i}{T}+\frac{X_i'\bar{M}_wF}{T}\gamma_i\right\|^2\leq\left\|\frac{X_i'\epsilon_i}{T}\right\|^2+\left\|\frac{X_i'\bar{M}_wF}{T}\right\|^2,$$

so that the order of magnitude of $\max_{1\leq i\leq n}\left\|\tilde{\beta}_i-\beta_i\right\|^2$ can be derived by studying $T^{-1}\max_{1\leq i\leq n}\left\|T^{-1/2}X_i'\epsilon_i\right\|^2$ and $\max_{1\leq i\leq n}\left\|T^{-1}X_i'\bar{M}_wF\right\|^2$. Consider the former. By the proof of Lemma A.1, we know that $E\left\|T^{-1/2}X_i'\epsilon_i\right\|^a$ is bounded by $E\left\|x_{it}\epsilon_{it}\right\|^a\leq E\left\|x_{it}\right\|^a\left|\epsilon_{it}\right|^a=E\left\|x_{it}\right\|^a E\left|\epsilon_{it}\right|^a$ by using Assumption 2*(iii)*. The largest $a$ for which this moment exists is $a=k/2$, whence $\max_{1\leq i\leq n}\left\|T^{-1/2}X_i'\epsilon_i\right\|^2=o_p\left(n^{2/k}\right)$. This entails $T^{-1}\max_{1\leq i\leq n}\left\|T^{-1/2}X_i'\epsilon_i\right\|^2=o_p\left(n^{2/k}T^{-1}\right)$. As far as $\max_{1\leq i\leq n}\left\|T^{-1}X_i'\bar{M}_wF\right\|^2$ is concerned, we know from the proof of Lemma A.1 that $\left\|T^{-1}X_i'\bar{M}_wF\right\|^2$ has magnitude $O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. When applying $\max_{1\leq i\leq n}$, this only affects the $x_{it}$s. To illustrate this, consider term $I$ in (40): $\frac{1}{\sqrt{nT}}\max_{1\leq i\leq n}\left\|\frac{X_i'FCD_w^{-1}\sqrt{n}\bar{\epsilon}'F}{T\sqrt{T}}\right\|^2\leq\frac{1}{\sqrt{nT}}\left(\max_{1\leq i\leq n}\left\|\frac{X_i}{\sqrt{T}}\right\|^2\right)\left\|\frac{F}{\sqrt{T}}\right\|^2\left\|\frac{\bar{\epsilon}'F}{\sqrt{nT}}\right\|^2$. We have $\left\|T^{-1/2}X_i\right\|^2=T^{-1}\sum_{t=1}^T x_{it}^2$, so that, based on (44), $\max_{1\leq i\leq n}\left\|T^{-1/2}X_i\right\|^2=o_p\left(n^{2/k}\right)$. Therefore, the whole expression is of order $o_p\left(n^{-1/2}T^{-1/2}n^{2/k}\right)$. Applying similar passages to terms $II$ and $III$ in (40) yields $\max_{1\leq i\leq n}\left\|T^{-1}X_i'\bar{M}_wF\right\|^2=o_p\left(n^{2/k-1/2}\delta_{nT}^{-1}\right)$. Part *(i)* follows putting everything together.

Consider part *(ii)*. The passages of the proof are rather repetitive. The main argument is that, based on (47), $\max_{1\leq t\leq T}\delta_{nt}^{-2}\left\|\hat{f}_t-H'f_t\right\|^2$ is bounded by terms such as $\delta_{nt}^{-2}\left\|\frac{1}{n}\sum_{j=1}^n\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\right\|^2\max_{1\leq t\leq T}\left\|x_{jt}\right\|^2$, etc. This entails that, when taking the maximum across $t$, the order of magnitude of the maximum is given by terms like $\max_{1\leq t\leq T}\left\|x_{jt}\right\|^2$, $\max_{1\leq t\leq T}\left\|f_t\right\|^2$ and $\max_{1\leq t\leq T}\left\|\epsilon_{jt}\right\|^2$, which are of order $o_p\left(T^{2/k}\right)$. This provides part *(ii)*.

The proof of part *(iii)* is based on (46):

$$\begin{aligned}
T\left\|\hat{\gamma}_i-\gamma_i\right\|^2 &\leq \left\|\left(\frac{\hat{F}'M_{Xi}\hat{F}}{T}\right)^{-1}\right\|_1^2 \left\|\frac{\hat{F}'M_{Xi}\epsilon_i}{\sqrt{T}}-\frac{\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma_i}{\sqrt{T}}\right\|^2 \\
&\leq \left\|\frac{\hat{F}'\epsilon_i}{\sqrt{T}}\right\|^2+\left\|\frac{\hat{F}'\left(\hat{F}-F\right)}{\sqrt{T}}\right\|^2\left\|\gamma_i\right\|^2.
\end{aligned}$$

By Assumption 3*(iii)*, $\max_{1\leq i\leq n}\left\|T^{-1/2}\hat{F}'\left(\hat{F}-F\right)\right\|^2\left\|\gamma_i\right\|^2$ has the same order of magnitude as $\left\|T^{-1/2}\hat{F}'\left(\hat{F}-F\right)\right\|^2$, i.e. $O_p\left(\sqrt{T}n^{-1}\right)+O_p\left(T^{-1/2}\right)$. As far as $\max_{1\leq i\leq n}\left\|T^{-1/2}\hat{F}'\epsilon_i\right\|^2$ is concerned

$$\max_{1\leq i\leq n}\left\|\frac{\hat{F}'\epsilon_i}{\sqrt{T}}\right\|^2\leq\max_{1\leq i\leq n}\left\|\frac{F'\epsilon_i}{\sqrt{T}}\right\|^2+T\max_{1\leq i\leq n}\left\|\frac{\left(\hat{F}-F\right)'\epsilon_i}{T}\right\|^2=I+II.$$

Consider $I$; based on the same arguments as in (39), we have $\max_{1\leq i\leq n}\left\|T^{-1/2}F'\epsilon_i\right\|^2\leq\max_{1\leq i\leq n}T^{-1}\sum_{t=1}^T\left\|f_t\epsilon_{it}\right\|^2$. Also, $E\left\|f_t\epsilon_{it}\right\|^a\leq E\left\|f_t\right\|^a E\left|\epsilon_{it}\right|^a<\infty$ with the largest $a$ being $a=2k$, whence $\max_{1\leq i\leq n}\left\|T^{-1/2}F'\epsilon_i\right\|^2=o_p\left(n^{2/k}\right)$. Turning to $II$, $\max_{1\leq i\leq n}\left\|T^{-1}\left(\hat{F}-F\right)'\epsilon_i\right\|^2$ can be studied using (41). It follows that $\max_{1\leq i\leq n}\left\|T^{-1}\left(\hat{F}-F\right)'\epsilon_i\right\|^2$ is bounded by the sum of terms like

$$\begin{aligned}
&\max_{1\leq i\leq n}\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^n\left(\frac{\epsilon_i'X_j}{\sqrt{T}}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\right\|^2 \\
\leq\ & \frac{1}{T}\max_{1\leq i\leq n}\left[\frac{1}{n}\sum_{j=1}^n\left\|\frac{\epsilon_i'X_j}{\sqrt{T}}\right\|^6\right]^{1/3}\left[\frac{1}{n}\sum_{j=1}^n\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{4/3}\left[\frac{1}{n}\sum_{j=1}^n\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/3} \\
\leq\ & \left[\frac{1}{nT}\sum_{j=1}^n\max_{1\leq i\leq n}\left\|\frac{\epsilon_i'X_j}{\sqrt{T}}\right\|^2\right]\left[\frac{1}{n}\sum_{j=1}^n\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{4/3}\left[\frac{1}{n}\sum_{j=1}^n\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/3},
\end{aligned}$$

which follows from (42) (first line) and from the $C_r$-inequality (second line). Note that $E\left\|T^{-1/2}\epsilon_i'X_j\right\|^a$ is bounded by $E\left\|\epsilon_{it}x_{jt}\right\|^a\leq E\left\|x_{jt}\right\|^a E\left|\epsilon_{it}\right|^a$ with $a=k$ at most. We have that $\max_{1\leq i\leq n}\left\|T^{-1/2}\epsilon_i'X_j\right\|^2=o_p\left(n^{2/k}\right)$. Applying the same logic to the squares of all the terms in (41), it follows that $II=o_p\left(n^{2/k}T\delta_{nT}^{-4}\right)$. Part *(iii)* follows from putting everything together.

Consider parts *(iv)* and *(v)*. Using the definition of $\hat{\epsilon}_{it}$:

$$\begin{aligned}
\max_{1\leq t\leq T}\hat{\epsilon}_{it}^2 \leq\ & \max_{1\leq t\leq T}\epsilon_{it}^2+\left\|\tilde{\beta}_i-\beta_i\right\|^2\max_{1\leq t\leq T}\left\|x_{it}\right\|^2+\left\|\hat{\gamma}_i-\gamma_i\right\|^2\max_{1\leq t\leq T}\left\|f_t\right\|^2 \\
&+\left\|\hat{\gamma}_i\right\|^2\max_{1\leq t\leq T}\left\|\hat{f}_t-f_t\right\|^2=I+II+III+IV.
\end{aligned}$$

Parts *(iv)* and *(v)* follow immediately from Assumptions 1 and 2. Explicit rates are derived using the other parts of this Lemma. Parts *(vi)* and *(vii)* can be proved similarly, using

$$\max_{1\leq i\leq n}\hat{\epsilon}_{it}^{2}\leq\max_{1\leq i\leq n}\epsilon_{it}^{2}+\max_{1\leq i\leq n}\left\|\tilde{\beta}_{i}-\beta_{i}\right\|^{2}\|x_{it}\|^{2}+\|f_{t}\|^{2}\max_{1\leq i\leq n}\|\hat{\gamma}_{i}-\gamma_{i}\|^{2}+\left\|\hat{f}_{t}-f_{t}\right\|^{2}\max_{1\leq i\leq n}\|\hat{\gamma}_{i}\|^{2},$$

and $\max_{1\leq i\leq n}\epsilon_{it}^{2}=o_{p}\left(n^{2/k}\right)$ by (44).

**Lemma A.7** *Let Assumptions 1-4 hold, and let $k$ denote the largest finite moment of $\epsilon_{it}$, $f_t$ and $x_{it}$:*

**A.7*(i)*** *if, in addition, Assumption 6 holds, then* $\left\|T^{-1}\hat{F}'\hat{F}-H\Sigma_{f}H'\right\|=O_{p}\left(T^{-1/2}\right)+O_{p}\left(n^{-1}\right)$;

**A.7*(ii)*** *if, in addition, Assumption 6 holds, then* $\max_{1\leq i\leq n}\left\|T^{-1}\hat{F}'M_{Xi}\hat{F}-\Sigma_{fM,i}\right\|=O_{p}\left(T^{-1/2}\right)+O_{p}\left(n^{-1}\right)$;

**A.7*(iii)*** *if, in addition, Assumption 8 holds, then* $\max_{1\leq t\leq T}\left\|\hat{\Sigma}_{\Gamma\epsilon,t}-H^{-1}\Sigma_{\Gamma\epsilon,t}\left(H^{-1}\right)'\right\|=o_{p}\left(T^{2/k}\delta_{nT}^{-1}\right)$;

**A.7*(iv)*** *if, in addition, Assumption 8 holds, then* $\max_{1\leq i\leq n}\left\|\hat{\Sigma}_{\gamma,i}-\Sigma_{\gamma,i}\right\|=o_{p}\left(\sqrt{T}n^{2/k}\delta_{nT}^{-2}\right)$;

**A.7*(v)*** *if, in addition, Assumption 6 holds, then* $\max_{1\leq i\leq n}\left\|T^{-1/2}F'M_{Xi}\epsilon_{i}-N_{i}\right\|=o_{p}\left(n^{1/k}T^{1/k-1/2}\right)$, *where $\{N_i\}_{i=1}^{n}$ is a sequence of i.i.d. Gaussian random variables, with variances $\Sigma_{fMe,i}$;*

**A.7*(vi)*** *if, in addition, Assumption 8 holds, then* $\max_{1\leq t\leq T}\left\|n^{-1/2}\sum_{i=1}^{n}\hat{\gamma}_{i}\epsilon_{it}-N_{t}\right\|=o_{p}\left(T^{1/k}n^{1/k-1/2}\right)$ $+\ o_{p}\left(T^{1/k}\delta_{nT}^{-1}\right)+\ o_{p}\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right)$, *where $\{N_t\}_{t=1}^{T}$ is a sequence of i.i.d. Gaussian random variables, with variances $\Sigma_{\Gamma\epsilon,t}$.*

**Proof.** As a preliminary result, note that Assumptions 1*(i)*, 2*(i)* and 6*(i)* entail that $\epsilon_{it}^2$, $f_t\epsilon_{it}$, $x_{it}\epsilon_{it}$, $vec\left(f_tf_t'\right)$, $vec\left(f_tx_{it}'\right)$ and $vec\left(f_tf_t'\epsilon_{it}^2\right)$ are all $L_{2+\delta}$-NED of size $\alpha'>\frac{1}{2}$ on $\{v_t\}_{t=-\infty}^{+\infty}$, for each $i$. These results are applications of Example 17.17 in Davidson (1994, p. 273), and are explicitly reported in Kao, Trapani and Urga (2012; see in particular Lemmas 8 and 9 therein). Similarly, Assumption 8 entails that $\epsilon_{it}^2$ is $L_{2+\delta}$-NED of size $\alpha'>\frac{1}{2}$ on $\{v_i\}_{i=-\infty}^{+\infty}$, for each $t$.

Consider part *(i)*, writing $T^{-1}\hat{F}'\hat{F}-\Sigma_f=\left(T^{-1}\hat{F}'\hat{F}-T^{-1}F'F\right)+\left(T^{-1}F'F-\Sigma_f\right)$. Lemma A.3 (parts *(ii)* and *(iii)*) entails that $T^{-1}\hat{F}'\hat{F}-T^{-1}F'F=O_p\left(\delta_{nT}^{-2}\right)$. The CLT for NED sequences can be applied (Theorem 24.6 and Corollary 24.7 in Davidson , 1994, p. 386-387), so that $E\left\|T^{-1}F'F-\Sigma_f\right\|^2$ $=E\left\|T^{-1}\sum_{t=1}^{T}\left(f_tf_t'-\Sigma_f\right)\right\|^2=O\left(T^{-1}\right)$. Putting all together, part *(i)* follows. As far as part *(ii)* is concerned, it follows immediately from noting that $\left\|T^{-1}\hat{F}'M_{Xi}\hat{F}-\Sigma_{fM,i}\right\|\leq\left\|T^{-1}\hat{F}'\hat{F}-\Sigma_f\right\|$ for each $i$, by definition of $M_{Xi}$.

As far as showing parts *(iii)-(vi)* is concerned, we extensively use the following result, which is an application of Theorem 2.1 in Berkes, Liu and Wu (2013; see also Theorem 2.2 in Ling, 2007). Given an $L_{2+\delta}$-NED zero mean sequence $Z_1, ..., Z_m$ of size (equal to or greater than) $\frac{1}{2}$, such that $E|Z_1|^k \leq M$ for some $k > 2$, and that the conditions spelt out in Assumptions 6*(ii)* and 6*(iii)* hold; and given a Brownian motion $W(\cdot)$ with $E\left[W^2(1)\right] = \lim_{m\to\infty} E\left[\left(m^{-1/2}\sum_{h=1}^{m} Z_h\right)^2\right]$, it holds that, redefining $Z_h$ in a richer probability space

$$\left\| \frac{1}{\sqrt{m}} \sum_{h=1}^{m} Z_h - W(1) \right\| = O_p\left(m^{1/k-1/2}\right). \tag{45}$$

Results like (45) are known as "Hungarian constructions"; see, *inter alia*, Csörgö and Révész (1975a,b), and Komlós, Major and Tusnády (1975, 1976); we also refer to Shorack and Wellner (1986) for a review. Hungarian constructions are usually stated in terms of the partial sum process $m^{-1/2}\sum_{h=1}^{\lfloor m\tau \rfloor} Z_h$ for $\tau \in [0,1]$; in that case, (45) is stated using the sup-norm. For our purposes, we only need to consider $\tau = 1$. The rate in (45) is sharp, and this result was shown, for the case of dependent data, only very recently (Berkes, Liu and Wu, 2013). By Assumptions 6 and 8, and by the fact that, as stated above, $\epsilon_{it}^2$, $f_t\epsilon_{it}$, $x_{it}\epsilon_{it}$, $vec(f_t f_t')$, $vec(f_t x_{it}')$ and $vec\left(f_t f_t' \epsilon_{it}^2\right)$ are all $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$, equation (45) can be applied to the normalized sums of all these sequences - in the case of $\epsilon_{it}^2$, to both sums across $t$ and across $i$. As a final remark, we point out that if all moments of $Z_h$ exist (e.g. if $Z_h$ is Gaussian), the rate in (45) becomes exponential, i.e. (45) holds with a rate $O_p\left(\frac{\ln m}{\sqrt{m}}\right)$.

We turn to the proof of part *(iii)* of the Lemma. We have

$$\begin{aligned}
\max_{1\leq t\leq T}\left\|\hat{\Sigma}_{\Gamma\epsilon,t} - \Sigma_{\Gamma\epsilon,t}\right\| &\leq \max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\gamma_i'\epsilon_{it}^2 - \Sigma_{\Gamma\epsilon,t}\right\| + \max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i\hat{\gamma}_i'\left(\hat{\epsilon}_{it}^2 - \epsilon_{it}^2\right)\right\| \\
&\quad + 2\max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\left(\hat{\gamma}_i - \gamma_i\right)'\hat{\epsilon}_{it}^2\right\| + \max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\gamma}_i - \gamma_i\right)\left(\hat{\gamma}_i - \gamma_i\right)'\hat{\epsilon}_{it}^2\right\| \\
&= I + II + III + IV.
\end{aligned}$$

Consider $I$. Recall that the sequence $z_{\epsilon\gamma,it} = \gamma_i\gamma_i'\epsilon_{it}^2 - \Sigma_{\Gamma\epsilon,t}$ is, as stated above, $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$. Therefore, by Theorem 17.5(b) in Davidson (1994, p. 264), $z_{\epsilon\gamma,it}$ is an $L_{2+\delta}$-mixingale of size $\min\left\{\alpha', \frac{k-1}{k-2}\right\} > \frac{1}{2}$. Using Assumption 3*(iii)* and Corollary 1 in Peligrad, Utev, and Wu (2007), it follows that $E\left\|n^{-1/2}\sum_{i=1}^{n} z_{\epsilon\gamma,it}\right\|^a \leq ME\left|\epsilon_{it}^2\right|^a < \infty$. By Assumption 1*(i)*, the largest $a$ for which $E\left\|n^{-1/2}\sum_{i=1}^{n} z_{\epsilon\gamma,it}\right\|^k < \infty$ is $k/2$. Thus, $\max_{1\leq t\leq T}\left\|n^{-1/2}\sum_{i=1}^{n} z_{\epsilon\gamma,it}\right\| = o_p\left(T^{2/k}\right)$, which entails $I = o_p\left(T^{2/k}n^{-1/2}\right)$. As far as $II$ is concerned, it has the same order of magnitude as $\max_{1\leq t\leq T}\left|\hat{\epsilon}_{it}^2 - \epsilon_{it}^2\right|$, given in Lemma A.6*(v)*. Turning to $III$, its order of magnitude is given by

$O_p\left(\|\hat{\gamma}_i-\gamma_i\|^2\right)\max_{1\leq t\leq T}\hat{\epsilon}_{it}^2$, which comes from Lemma A.6*(iv)*. Term $IV$ is dominated. Putting all together, part *(iii)* of the Lemma follows.

As far as part *(iv)* is concerned, the proof is similar, in spirit, to that of part *(iii)*. Recall that $\hat{\Sigma}_{\gamma i}=(Q_i')^{-1}D_{0,i}(Q_i)^{-1}$; the rates for $Q_i=T^{-1}\hat{F}'M_{Xi}\hat{F}$ are given by part *(ii)* of this Lemma. Based on the definition of $D_{0,i}$, we have

$$\begin{aligned}
& \max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t\hat{f}_t'\hat{\epsilon}_{it}^2-H'\Sigma_f HE\left(\epsilon_{it}^2\right)\right\| \\
\leq\; & \max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}f_tf_t'\epsilon_{it}^2-H'\Sigma_f HE\left(\epsilon_{it}^2\right)\right\|+\max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t\hat{f}_t'\left(\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right)\right\| \\
& +2\max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}f_t\left(\hat{f}_t-f_t\right)'\hat{\epsilon}_{it}^2\right\|+\max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{f}_t-\hat{f}_t\right)\left(\hat{f}_t-\hat{f}_t\right)'\hat{\epsilon}_{it}^2\right\| \\
=\; & I+II+III+IV.
\end{aligned}$$

As far as $I$ is concerned, the proof is similar to that of part *(iii)* of this Lemma, upon recalling that $f_tf_t'\epsilon_{it}^2$ is, across $t$, $L_{2+\delta}$-NED of size $\alpha'>\frac{1}{2}$. Indeed, the largest $a$ for which $E\left\|f_tf_t'\epsilon_{it}^2\right\|^a\leq E\left\|f_t\right\|^{2a}E\left|\epsilon_{it}\right|^{2a}$ is $a=k/2$, and that $T^{-1}\sum_{t=1}^{T}f_tf_t'\epsilon_{it}^2-H'\Sigma_f HE\left(\epsilon_{it}^2\right)=O_p\left(T^{-1/2}\right)$; hence, $I=o_p\left(n^{2/k}T^{-1/2}\right)$. Considering $II$, we have

$$\begin{aligned}
II &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_t\hat{f}_t'\right\|\max_{1\leq i\leq n}\left|\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right| \\
&\leq \left[\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_t\hat{f}_t'\right\|^2\right]^{1/2}\left[E\left(\max_{1\leq i\leq n}\left|\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right|\right)^2\right]^{1/2};
\end{aligned}$$

applying Lemma A.6*(v)*, we have $II=o_p\left(T^{-2/k}\delta_{nT}^{-2}\right)$ . Considering $III$, a similar logic as above yields

$$\begin{aligned}
III &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|f_t\left(\hat{f}_t-f_t\right)'\right\|\max_{1\leq i\leq n}\hat{\epsilon}_{it}^2 \\
&\leq \left[\frac{1}{T}\sum_{t=1}^{T}\left\|f_t\left(\hat{f}_t-f_t\right)'\right\|^2\right]^{1/2}\left[E\left(\max_{1\leq i\leq n}\hat{\epsilon}_{it}^2\right)^2\right]^{1/2};
\end{aligned}$$

also,

$$\left[\frac{1}{T}\sum_{t=1}^{T}\left\|f_t\left(\hat{f}_t-f_t\right)'\right\|^2\right]^{1/2}\leq\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\left\|f_t\left(\hat{f}_t-f_t\right)'\right\|=O_p\left(\sqrt{T}\delta_{nT}^{-2}\right),$$

by the $C_r$-inequality and Lemma A.3*(ii)*. Using Lemma A.6*(vi)*, we have $III=o_p\left(n^{2/k}T^{1/2}\delta_{nT}^{-2}\right)+o_p\left(n^{4/k}T^{1/2}\phi_{nT}^{-2}\delta_{nT}^{-2}\right)+o_p\left(n^{2/k-2}T^{3/2}\delta_{nT}^{-2}\right)$. Term $IV$ is dominated. Putting all together, part *(iv)* follows.

Consider now part *(v)*. We have $T^{-1/2}F'M_{Xi}\epsilon_i = T^{-1/2}\sum_{t=1}^{T} f_t\epsilon_{it} - \left(T^{-1}\sum_{t=1}^{T} f_t x_{it}'\right)\left(T^{-1}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1}$ $T^{-1/2}\sum_{t=1}^{T} x_{it}\epsilon_{it}$. Since $f_t\epsilon_{it}$ and $x_{it}\epsilon_{it}$ are $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$, equation (45) holds with $k^* = 4$: there are two sequences of *i.i.d.* Gaussian, zero mean random variables, say $\left\{N_{it}^{f\epsilon}\right\}_{t=1}^{T}$ and $\{N_{it}^{x\epsilon}\}_{t=1}^{T}$, such that $E\left[\left(N_{it}^{f\epsilon}\right)^2\right] = \Sigma_{f\epsilon,i}$ and $E\left[(N_{it}^{x\epsilon})^2\right] = \Sigma_{x\epsilon,i}$ and $T^{1/k-1/2}\left\|T^{-1/2}\sum_{t=1}^{T} f_t\epsilon_{it} - T^{-1/2}\sum_{t=1}^{T} N_{it}^{f\epsilon}\right\| = O_p(1)$ and $T^{1/k-1/2}\left\|T^{-1/2}\sum_{t=1}^{T} x_{it}\epsilon_{it} - T^{-1/2}\sum_{t=1}^{T} N_{it}^{x\epsilon}\right\| = O_p(1)$. Further, by the CLT for NED processes (see e.g. Theorem 24.6 in Davidson , 1994, p. 386), $T^{-1}\sum_{t=1}^{T} f_t x_{it}' = \Sigma_{fx,i} + O_p\left(T^{-1/2}\right)$ and $T^{-1}\sum_{t=1}^{T} x_{it}x_{it}' = \Sigma_{xx,i} + O_p\left(T^{-1/2}\right)$. Putting all together, and defining the *i.i.d.* Gaussian sequence $N_i \equiv T^{-1/2}\sum_{t=1}^{T} N_{it}^{f\epsilon} - \Sigma_{fx,i}\Sigma_{xx,i}^{-1}T^{-1/2}\sum_{t=1}^{T} N_{it}^{x\epsilon}$, we have that $T^{1/k-1/2}$ $\left\|T^{-1/2}F'M_{Xi}\epsilon_i - N_i\right\| = O_p(1)$. Note further that, $E\left\|T^{-1/2}F'M_{Xi}\epsilon_i\right\|^r \leq E\left\|T^{-1/2}\sum_{t=1}^{T} f_t\epsilon_{it}\right\|^r$; note that $f_t\epsilon_{it}$ is $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$ on $\{v_t\}_{t=-\infty}^{+\infty}$, which entails that it is an $L_{2+\delta}$-mixingale of size $\min\left\{\alpha', \frac{k-1}{k-2}\right\} > \frac{1}{2}$. Therefore, using again Corollary 1 in Peligrad, Utev, and Wu (2007), $E\left\|T^{-1/2}\sum_{t=1}^{T} f_t\epsilon_{it}\right\|^r \leq M\ E\left\|f_t\epsilon_{it}\right\|^r$; by Assumptions 1*(i)*, 2*(i)* and 2*(iii)*, the largest $r$ for which $E\left\|T^{-1/2}F'M_{Xi}\epsilon_i\right\|^r < \infty$ is $r = k$. Thus, $\max_{1\leq i\leq n} T^{1/k-1/2}\left\|T^{-1/2}F'M_{Xi}\epsilon_i - N_i\right\| = o_p\left(n^{1/k}\right)$, which proves part *(v)*.

The proof of part *(vi)* is similar. Indeed, given an independent, zero mean, Gaussian sequence $\{N_{it}^n\}_{i=1}^n$ with $E\left(N_{it}^n\right)^2 = \gamma_i\gamma_i'\ E\left(\epsilon_{it}^2\right)$, write

$$
\begin{aligned}
& \max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\gamma}_i\epsilon_{it} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} N_{it}^n\right\| \\
\leq\ & \max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_i\epsilon_{it} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} N_{it}^n\right\| + \max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{\gamma}_i - \gamma_i\right)\epsilon_{it}\right\| \\
=\ & I + II.
\end{aligned}
$$

By virtue of Assumption 8, and using similar considerations as for the proof of part *(v)*, term $I$ satisfies (45) with

$$
n^{1/2-1/k}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_i\epsilon_{it} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} N_{it}^n\right\| = O_p(1);
$$

also, the largest existing moment over $t$ is of order $k$. Thus, $\max_{1\leq t\leq T} n^{1/k-1/2}\left\|n^{-1/2}\sum_{i=1}^{n}\gamma_i\epsilon_{it} - n^{-1/2}\sum_{i=1}^{n} N_{it}^n\right\| = o_p\left(T^{1/k}\right)$, whence $I = o_p\left(T^{1/k}n^{1/k-1/2}\right)$. Consider now $II$. By (46), we can write

$$
II = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(F'M_{Xi}F\right)^{-1}\left(F'M_{Xi}\epsilon_i\right)\epsilon_{it} + II_b = II_a + II_b,
$$

where $II_b$ contains the remainder of $\left(\hat{\gamma}_i - \gamma_i\right)$. By the results in the proof of Theorem 1, and by using the Cauchy-Schwartz inequality, it follows immediately that $II_b = o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right) + o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right)$. Also, $II_a \leq M\ n^{-1/2}\ T^{-1}\ \left\|\sum_{i=1}^{n}\sum_{s=1}^{T} f_s\epsilon_{is}\epsilon_{it}\right\|$, so that

$$II_a \leq M\frac{1}{\sqrt{n}T}\left\|\sum_{i=1}^{n}\sum_{s=1}^{T} f_s\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\right\| + M\frac{1}{\sqrt{n}T}\left\|\sum_{i=1}^{n}\sum_{s=1}^{T} f_s E\left(\epsilon_{is}\epsilon_{it}\right)\right\| = II_{a,1} + II_{a,2}.$$

We have that $II_{a,1}$ is bounded by the square root of

$$\begin{aligned}
& E\left\|\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{u=1}^{T}\sum_{s=1}^{T} f_s f_u'\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\left[\epsilon_{ju}\epsilon_{jt} - E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\| \\
\leq\ & \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{u=1}^{T}\sum_{s=1}^{T} E\left\|f_s f_u'\right\| E\left|\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\left[\epsilon_{ju}\epsilon_{jt} - E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right| \\
\leq\ & M\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{u=1}^{T}\sum_{s=1}^{T} E\left|\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\left[\epsilon_{ju}\epsilon_{jt} - E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right| \\
\leq\ & M'\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{u=1}^{T}\sum_{s=1}^{T} E\left|\epsilon_{is}\epsilon_{it}\epsilon_{ju}\epsilon_{jt}\right| \leq M''\frac{1}{T},
\end{aligned}$$

on account of Assumptions 2*(iii)* and 1*(iii)*(c); hence, $II_{a,1} = O_p\left(T^{-1/2}\right)$. The same logic entails $II_{a,2} = O_p\left(T^{-1/2}\right)$ also. Putting all together, $II = o_p\left(T^{1/k}\delta_{nT}^{-1}\right) + o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right)$. Defining $n^{-1/2}\sum_{i=1}^{n} N_{it}^{n} = N_t$, part *(vi)* follows. QED

# Appendix B: Proofs

Similarly to Appendix A, in this section we set the rotation matrix $H = I_r$ whenever possible in order to simplify the notation.

**Proof of Theorem 1**. By definition, we have

$$\sqrt{T}\left(\hat{\gamma}_i - \gamma_i\right) = \left(\frac{\hat{F}' M_{Xi} \hat{F}}{T}\right)^{-1} \left[\frac{\hat{F}' M_{Xi} \epsilon_i}{\sqrt{T}} - \frac{\hat{F}' M_{Xi}\left(\hat{F} - F\right)\gamma_i}{\sqrt{T}}\right]. \tag{46}$$

We start by considering the denominator of (46):

$$\begin{aligned} & \frac{\hat{F}' M_{Xi} \hat{F}}{T} - \frac{F' M_{Xi} F}{T} \\ = \quad & \frac{F' M_{Xi}\left(\hat{F} - F\right)}{T} + \frac{\left(\hat{F} - F\right)' M_{Xi} F}{T} - \frac{\left(\hat{F} - F\right)' M_{Xi}\left(\hat{F} - F\right)}{T} = I + I' - II. \end{aligned}$$

Repeated application of Lemma A.3 yields $I = O_p\left(\delta_{nT}^{-2}\right)$ and $II = O_p\left(\delta_{nT}^{-4}\right)$. Thus, as $(n, T) \to \infty$, $T^{-1}\hat{F}' M_{Xi} \hat{F} = T^{-1} F' M_{Xi} F + o_p(1)$.

We turn to the numerator of (46). It holds that

$$\frac{\hat{F}' M_{Xi} \epsilon_i}{\sqrt{T}} = \frac{F' M_{Xi} \epsilon_i}{\sqrt{T}} + \frac{\left(\hat{F} - F\right)' M_{Xi} \epsilon_i}{\sqrt{T}} = I + II.$$

By applying a similar logic as in the proof of Lemma A.4, it can be shown that $I = O_p(1)$. As far as $II$ is concerned, note

$$II = \sqrt{T}\frac{\left(\hat{F} - F\right)' \epsilon_i}{T} + \frac{\left(\hat{F} - F\right)' X_i}{T}\left(\frac{X_i' X_i}{T}\right)^{-1} \frac{X_i' \epsilon_i}{\sqrt{T}};$$

applying Lemma A.2*(i)* (to the first term), and Lemma A.3*(i)* and Assumptions 2*(i)* and 1*(i)* (to the second term), it follows that $II = O_p\left(\sqrt{T}\delta_{nT}^{-2}\right)$. Thus, the numerator of (46) is of order $O_p(1) + O_p\left(\frac{\sqrt{T}}{n}\right)$.

Finally, as $(n, T) \to \infty$ under the restriction $\frac{\sqrt{T}}{n} \to 0$, (46) becomes

$$\sqrt{T}\left(\hat{\gamma}_i - \gamma_i\right) = \left(\frac{F' M_{Xi} F}{T}\right)^{-1} \frac{F' M_{Xi} \epsilon_i}{\sqrt{T}} + o_p(1);$$

equation (6) follows from Assumption 5*(i)*. QED

**Proof of Theorem 2.** Using (37), we can write

$$\begin{aligned}
\hat{f}_t - f_t &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)' x_{jt} \\
&\quad -\frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)' x_{jt} - \frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j-\beta_j\right)' x_{jt} \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\gamma_j' f_t - \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\epsilon_{jt} \\
&\quad +\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\gamma_j' f_t + \frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\epsilon_{jt} + \frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\epsilon_{jt} \\
&= I - II - III - IV - V + VI + VII + VIII.
\end{aligned} \tag{47}$$

The order of magnitude of $I$ follows exactly from the same passages as in the proof of Lemma A.5, with $I = O_p\left(\phi_{nT}^{-2}\right)$. Consider $II$; omitting $\gamma_j$ in view of Assumption 3*(iii)*, we have

$$II = \frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\Upsilon_j' x_{jt} + \frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\bar{\Upsilon}_j' x_{jt} = II_a + II_b;$$

we have shown that $II_a = O_p\left(n^{-1/2}T^{-1/2}\right)$ and $II_b = O_p\left(n^{-1/2}T^{-1/2}\right) + O_p\left(n^{-1}\right)$ in the proof of Lemma A.3, so that $II = O_p\left(n^{-1/2}T^{-1/2}\right) + O_p\left(n^{-1}\right)$. Using Lemma A.3*(i)*, it can be shown that $III = O_p\left(\phi_{nT}^{-2}\right)$. As far as $IV$ is concerned, note that

$$IV = \frac{\hat{F}'}{\sqrt{T}}\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\Upsilon_j\gamma_j' f_t + \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\bar{\Upsilon}_j\gamma_j' f_t = IV_a + IV_b$$

Similar passages as in the proof of the order of magnitude of $II_a$, and the fact that $E\left\|f_t\right\| \leq M$ entail $IV_a = O_p\left(n^{-1/2}T^{-1/2}\right)$. Similarly, $IV_b$ is bounded by $\left\|f_t\right\|\left[E\left\|\frac{\hat{F}'X_j}{T}\right\|^2\right]^{1/2}\left[E\left\|\bar{\Upsilon}_j\right\|^2\right]^{1/2}$, which is $O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$ using Lemma A.1. Thus, $IV = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Turning to $V$, we have

$$\begin{aligned}
V &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\epsilon_{jt} \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\epsilon_{jt} \\
&\quad +\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_wF}{T}\gamma_j\right)\epsilon_{jt} = V_a + V_b.
\end{aligned}$$

We start from $V_b \leq n^{-1}\sum_{j=1}^{n}\left\|\frac{\hat{F}'X_j}{T}\right\|\left\|\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\right\|\left\|\frac{X_j'\bar{M}_wF}{T}\right\|\left\|\gamma_j\right\|\left|\epsilon_{jt}\right|$. Using Assumptions

3*(iii)* and 4*(i)*, $V_b$ is bounded by $E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|\left\|\frac{X_j'\bar{M}_wF}{T}\right\||\epsilon_{jt}|\right] \leq \left(E\left\|\frac{\hat{F}'X_j}{T}\right\|^6\right)^{1/6}\left(E\left\|\frac{X_j'\bar{M}_wF}{T}\right\|^{3/2}\right)^{2/3}$ $\left(E|\epsilon_{jt}|^6\right)^{1/6} = O\left(n^{-1}\right) + O\left(n^{-1/2}T^{-1/2}\right)$, where the passage in the middle follows from Holder's inequality. Consider now $V_a$:

$$
\begin{aligned}
V_a &= \frac{1}{n}\sum_{j=1}^n\left(\frac{F'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\epsilon_{jt} \\
&+\frac{1}{n}\sum_{j=1}^n\frac{\left(\hat{F}-F\right)'X_j}{T}\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\epsilon_{jt} = V_{a,1}+V_{a,2}.
\end{aligned}
$$

Consider $V_{a,2}$:

$$
\begin{aligned}
V_{a,2} &\leq \frac{1}{n}\sum_{j=1}^n\left\|\frac{\left(\hat{F}-F\right)'X_j}{T}\right\|\left\|\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\right\|\left\|\frac{X_j'\bar{M}_w\epsilon_j}{T}\right\||\epsilon_{jt}| \\
&\leq M\frac{1}{n}\sum_{j=1}^n\left\|\frac{\left(\hat{F}-F\right)'X_j}{T}\right\|\left\|\frac{X_j'\bar{M}_w\epsilon_j}{T}\right\||\epsilon_{jt}|,
\end{aligned}
$$

using Assumption 4*(i)*. Further, $E\left[\left\|\frac{\left(\hat{F}-F\right)'X_j}{T}\right\|\left\|\frac{X_j'\bar{M}_w\epsilon_j}{T}\right\||\epsilon_{jt}|\right] \leq \left(E\left\|\frac{\left(\hat{F}-F\right)'X_j}{T}\right\|^{3/2}\right)^{2/3}\left(E\left\|\frac{X_j'\bar{M}_w\epsilon_j}{T}\right\|^6\right)^{1/6}$ $\left(E|\epsilon_{jt}|^6\right)^{1/6}$, again by Holder's inequality. Using Lemma A.3(i), Assumption 2*(iv)* and similar passages as in the proof of (39), and Assumption 1*(i)*, we have $V_{a,2} = O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$. Turning to $V_{a,1}$

$$
\begin{aligned}
V_{a,1} &= \frac{1}{n}\sum_{j=1}^n\left(\frac{F'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\frac{X_j'\bar{M}_wE\left(\epsilon_j\epsilon_{jt}\right)}{T} \\
&+\frac{1}{n}\sum_{j=1}^n\left(\frac{F'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\frac{X_j'\bar{M}_w\left[\epsilon_j\epsilon_{jt}-E\left(\epsilon_j\epsilon_{jt}\right)\right]}{T} = V_{a,1,1}+V_{a,1,2}.
\end{aligned}
$$

By virtue of Assumption 4*(i)*, $V_{a,1,1} \leq M\ n^{-1}\ T^{-2}\ \sum_{j=1}^n\ \|F'X_j\|\ \left\|X_j'\bar{M}_wE\left(\epsilon_j\epsilon_{jt}\right)\right\|$. We have $E\left[\left\|\frac{F'X_j}{T}\right\|\left\|\frac{X_j'\bar{M}_wE(\epsilon_j\epsilon_{jt})}{T}\right\|\right] \leq \left(E\left\|\frac{F'X_j}{T}\right\|^2\right)^{1/2}\left(E\left\|\frac{X_j'\bar{M}_wE(\epsilon_j\epsilon_{jt})}{T}\right\|^2\right)^{1/2}$, with $E\left\|\frac{F'X_j}{T}\right\|^2 \leq M$ by Assumption 2*(i)*. Further,

$$
\begin{aligned}
E\left\|\frac{X_j'\bar{M}_wE\left(\epsilon_j\epsilon_{jt}\right)}{T}\right\|^2 &\leq \frac{1}{T^2}\sum_{s=1}^T\sum_{u=1}^T E\left[\|x_{js}\|\|x_{ju}\|\right]E\left(\epsilon_{js}\epsilon_{jt}\right)E\left(\epsilon_{ju}\epsilon_{jt}\right) \\
&\leq M\frac{1}{T^2}\left[\sum_{s=1}^T E\left(\epsilon_{js}\epsilon_{jt}\right)\right]^2 = O\left(\frac{1}{T^2}\right),
\end{aligned}
$$

where we have used Assumptions 4*(i)*, 2*(i)* and 1*(ii)*(a). Consider now $V_{a,1,2}$; this is bounded by the square root of

$$E\left\{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{F'X_j}{T}\right)\left(\frac{F'X_k}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_k'\bar{M}_wX_k}{T}\right)^{-1}\right.$$
$$\left.\times\frac{X_j'\bar{M}_w\left[\epsilon_j\epsilon_{jt}-E\left(\epsilon_j\epsilon_{jt}\right)\right]}{T}\frac{X_k'\bar{M}_w\left[\epsilon_k\epsilon_{kt}-E\left(\epsilon_k\epsilon_{kt}\right)\right]}{T}\right\};$$

after some algebra, this is bounded by

$$\begin{aligned}
& E\left\{\frac{1}{n^2T^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{F'X_j}{T}\right)\left(\frac{F'X_k}{T}\right)\sum_{s=1}^{T}\sum_{u=1}^{T}x_{js}x_{ku}\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\left[\epsilon_{ju}\epsilon_{jt}-E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\}\\
= & \frac{1}{n^2T^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left[\left(\frac{F'X_j}{T}\right)\left(\frac{F'X_k}{T}\right)x_{js}x_{ku}\right]E\left\{\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\left[\epsilon_{ju}\epsilon_{jt}-E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\}\\
\leq & \frac{1}{n^2T^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left\{\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\left[\epsilon_{ju}\epsilon_{jt}-E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\}\\
\leq & \frac{1}{nT}E\left|\frac{1}{\sqrt{nT}}\sum_{j=1}^{n}\sum_{s=1}^{T}\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\right|^2,
\end{aligned}$$

by using Assumption 2*(iii)* in the second line, Assumption 2*(i)* in the third line, and Assumption 1*(iii)*(c) in the final passage. Thus, $V_{a,1,2}=O_p\left(n^{-1/2}T^{-1/2}\right)$. Putting all together, $V=O_p\left(T^{-1}\right)$ $+\,O_p\left(n^{-1/2}T^{-1/2}\right)$. The proofs of $VI=O_p\left(n^{-1/2}T^{-1/2}\right)$, $VII=O_p\left(n^{-1/2}\right)$ and $VIII=O_p\left(\delta_{nT}^{-2}\right)$ are based on the same arguments as in Bai (2003), since the estimation error $\tilde{\beta}_j-\beta_j$ does not appear in their expression. Putting everything together, as $(n,T)\to\infty$ with $\frac{\sqrt{n}}{T}\to 0$, the term that dominates in the expansion of $\hat{f}_t-f_t$ is $VII$, whose asymptotics is exactly the same as studied in Bai (2003, Theorem 1). QED

**Proof of Theorem 3.** Prior to proving the Theorem, we lay out some preliminary results and notation. We write

$$\hat{\gamma}_i-\widehat{\bar{\gamma}}=\left(\gamma_i-\bar{\gamma}\right)+\left(\hat{\gamma}_i-\gamma_i\right)-\left(\widehat{\bar{\gamma}}-\bar{\gamma}\right)=a_i+b_i-c_i.$$

Under $H_0^a$, $a_i = 0$; also, $b_i$ can be rewritten as $b_i = \hat{\gamma}_i - \bar{\gamma}$. Using (46), we have

$$\begin{aligned} b_i &= \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} F' M_{Xi} \epsilon_i + \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} \left(\hat{F} - F\right)' M_{Xi} \epsilon_i \\ &\quad - \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} \hat{F}' M_{Xi} \left(\hat{F} - F\right) \gamma_i \\ &= b_{1i} + b_{2i}, \end{aligned} \tag{48}$$

where we define $b_{1i} = \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} F' M_{Xi} \epsilon_i$ and $b_{2i}$ is the remainder. Further, we can write $\hat{\Sigma}_{\gamma i}^{-1} = \Sigma_{\gamma i}^{-1} - \Sigma_{\gamma i}^{-1} \left(\hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i}\right) \Sigma_{\gamma i}^{-1} + o_p \left(\left\| \hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i} \right\|\right)$ for each $i$. Neglecting higher order terms that depend on $o_p \left(\left\| \hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i} \right\|\right)$, we have

$$\begin{aligned} & T \left(\hat{\gamma}_i - \widehat{\bar{\gamma}}\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\hat{\gamma}_i - \widehat{\bar{\gamma}}\right) \\ = \; & T \left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right) + T b_{1i}' \Sigma_{\gamma i}^{-1} \left(\hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i}\right) \Sigma_{\gamma i}^{-1} b_{1i} + T b_{2i}' \hat{\Sigma}_{\gamma i}^{-1} b_{2i} \\ & + 2 T b_{1i}' \hat{\Sigma}_{\gamma i}^{-1} b_{2i} + T \left(\widehat{\bar{\gamma}} - \bar{\gamma}\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\widehat{\bar{\gamma}} - \bar{\gamma}\right) - 2T \left(\widehat{\bar{\gamma}} - \bar{\gamma}\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\hat{\gamma}_i - \bar{\gamma}\right) \\ = \; & T \left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right) + I_i + II_i + III_i + IV_i - V_i. \end{aligned} \tag{49}$$

After this preliminary calculations, we turn to proving (20). In order to do this, we firstly show that $\max_{1 \le i \le n} T \left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right)$ can be approximated by the maximum of a sequence of independent random variables with a $\chi_r^2$ distribution, up to a negligible error. Given that the maximum of a sequence of chi-squares is of order $O_p(\ln n)$, the approximation error should be $o_p(\ln n)$ at most. Secondly, we show that $I_i - V_i$ in (49) are also all $o_p(\ln n)$ uniformly in $i$.

Consider $\max_{1 \le i \le n} T \left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right)$, and consider in particular the sequence $\left\{\sqrt{T} b_{1i}\right\}_{i=1}^n$. It holds that $\sqrt{T} b_{1i} = \left[T^{-1} \hat{F}' M_{Xi} \hat{F}\right]^{-1} \left[T^{-1/2} F' M_{Xi} \epsilon_i\right]$. As far as the numerator of this expression is concerned, by Lemma A.7*(v)* we write $T^{-1/2} F' M_{Xi} \epsilon_i = N_i + R_{Ni}$ with $N_i$ defined in Lemma A.7 as being zero mean Gaussian with covariance matrix $\Sigma_{fMe,i}$, and $R_{Ni} = o_p\left(n^{1/k_1} T^{1/k_1 - 1/2}\right)$. As far as the denominator of $\sqrt{T} b_{1i}$ is concerned, based on Lemma A.7*(ii)* we write $\left[T^{-1} \hat{F}' M_{Xi} \hat{F}\right]^{-1} = \Sigma_{fM,i}^{-1} + R_{\Sigma fM,i}$ with $R_{\Sigma fM,i} = O_p\left(T^{-1/2}\right) + O_p\left(n^{-1}\right)$. Hence we write

$$\sqrt{T} b_{1i} = \left[\Sigma_{fM,i}^{-1} + R_{\Sigma fM,i}\right] \left[N_i + R_{Ni}\right]. \tag{50}$$

Based on (50), and on the definitions of $\Sigma_{fMe,i}$ and of $\Sigma_{fM,i}$, it holds that

$$
\begin{aligned}
T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right) &= N_i'\Sigma_{fMe,i}^{-1}N_i+2N_i'\Sigma_{fM,i}^{-1}\Sigma_{\gamma i}^{-1}R_{Ni}+2R_{Ni}'\Sigma_{fMe,i}^{-1}N_i \qquad (51)\\
&+2N_i'\Sigma_{fM,i}^{-1}\Sigma_{\gamma i}^{-1}R_{\Sigma fM,i}N_i+2N_i'\Sigma_{fM,i}^{-1}\Sigma_{\gamma i}^{-1}R_{\Sigma fM,i}R_{Ni}\\
&+R_{Ni}'\Sigma_{fMe,i}^{-1}R_{Ni}+2R_{Ni}'\Sigma_{fM,i}^{-1}\Sigma_{\gamma i}^{-1}R_{\Sigma fM,i}R_{Ni}\\
&+N_i'R_{\Sigma fM,i}\Sigma_{\gamma i}^{-1}R_{\Sigma fM,i}N_i+2N_i'R_{\Sigma fM,i}\Sigma_{\gamma i}^{-1}R_{\Sigma fM,i}R_{Ni}\\
&+R_{Ni}'R_{\Sigma fM,i}\Sigma_{\gamma i}^{-1}R_{\Sigma fM,i}R_{Ni}\\
&= N_i'\Sigma_{fMe,i}^{-1}N_i+I_i^{b1}+II_i^{b1}+III_i^{b1}+IV_i^{b1}+V_i^{b1}+VI_i^{b1}\\
&+VII_i^{b1}+VIII_i^{b1}+IX_i^{b1}.
\end{aligned}
$$

We note that the distribution of $N_i'\Sigma_{fMe,i}^{-1}N_i$ is $\chi_r^2$. We now show that, in (51), $\max_{1\le i\le n}I_i^{b1}$, ..., $\max_{1\le i\le n}IX_i^{b1}$ are all $o_p(1)$. Consider $\max_{1\le i\le n}I_i^{b1}$; this is bounded by $\max_{1\le i\le n}\|N_i\|\max_{1\le i\le n}\|R_{Ni}\|=o_p\left(n^{1/k_1}T^{1/k_1-1/2}\sqrt{\ln n}\right)$, in view of Lemma A.7*(v)* and the fact that $\max_{1\le i\le n}\|N_i\|=O_p\left(\sqrt{\ln n}\right)$. The same holds for $\max_{1\le i\le n}II_i^{b1}$. Turning to $\max_{1\le i\le n}III_i^{b1}$, it is bounded by $\max_{1\le i\le n}\|N_i\|^2\max_{1\le i\le n}\|R_{\Sigma fM,i}\|$ $=O_p\left(T^{-1/2}\ln n\right)+O_p\left(n^{-1}\ln n\right)$ by virtue of Lemma A.7*(ii)*. As far as $\max_{1\le i\le n}IV_i^{b1}$ is concerned, it is bounded by $\max_{1\le i\le n}\|N_i\|\max_{1\le i\le n}\|R_{\Sigma fM,i}\|\max_{1\le i\le n}\|R_{Ni}\|$, and therefore it is dominated by the previously analyzed terms. Also, $\max_{1\le i\le n}V_i^{b1}$ has the same order of magnitude as $\max_{1\le i\le n}\|R_{Ni}\|^2$, thereby being dominated by the other terms. Similarly, $\max_{1\le i\le n}VI_i^{b1}$ is bounded by $\max_{1\le i\le n}\|R_{Ni}\|^2\max_{1\le i\le n}\|R_{\Sigma fM,i}\|$, and therefore it is also dominated. Turning to $\max_{1\le i\le n}VII_i^{b1}$, it is bounded by $\max_{1\le i\le n}\|N_i\|^2\max_{1\le i\le n}\|R_{\Sigma fM,i}\|^2$, so that it is smaller than $\max_{1\le i\le n}III_i^{b1}$, and therefore negligible. Similarly, $\max_{1\le i\le n}VIII_i^{b1}$ is bounded by $\max_{1\le i\le n}\|N_i\|$ $\max_{1\le i\le n}\|R_{\Sigma fM,i}\|^2\max_{1\le i\le n}\|R_{Ni}\|$, which is dominated by $\max_{1\le i\le n}IV_i^{b1}$, and thus negligible. Finally, $\max_{1\le i\le n}IX_i^{b1}$ is bounded by $\max_{1\le i\le n}\|R_{\Sigma fM,i}\|^2\max_{1\le i\le n}\|R_{Ni}\|^2$, and it is dominated. Therefore

$$
\max_{1\le i\le n}T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right)=\max_{1\le i\le n}N_i'\Sigma_{fMe,i}^{-1}N_i+o_p\left[(nT)^{1/k_1}\sqrt{\frac{\ln n}{T}}\right]+O_p\left(\frac{\ln n}{\sqrt{T}}\right)+O_p\left(\frac{\ln n}{n}\right). \qquad (52)
$$

After proving that $\max_{1\le i\le n}T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right)$ can be approximated by $\max_{1\le i\le n}N_i'\Sigma_{fMe,i}^{-1}N_i$, we turn again to equation (49). We now show that $\max_{1\le i\le n}I_i$, ..., $\max_{1\le i\le n}V_i$ are all $o_p(\ln n)$. Consider $I_i$; it holds that

$$
\max_{1\le i\le n}I_i\le\left\|\max_{1\le i\le n}T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right)\right\|\left\|\max_{1\le i\le n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|.
$$

Equation (52) implies that $\max_{1\leq i\leq n} T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right) = O_p(\ln n)$; thus, applying Lemma A.7*(iv)*, $\max_{1\leq i\leq n} I_i = o_p\left(\sqrt{T}n^{2/k_1}\delta_{nT}^{-2}\ln n\right)$. Turning to $\max_{1\leq i\leq n} II_i$, note that, in equation (48), $b_{2i}$ is defined as

$$b_{2i} = \left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\left(\hat{F}-F\right)'M_{Xi}\epsilon_i - \left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma_i;$$

further, by the invertibility of $\Sigma_{\gamma i}^{-1}$ and Lemma A.7*(iv)*, $\max_{1\leq i\leq n} T\left(b_{2i}'\hat{\Sigma}_{\gamma i}^{-1}b_{2i}\right)$ has the same order of magnitude as $\max_{1\leq i\leq n}\left\|\sqrt{T}b_{2i}\right\|^2 \left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|$. Considering $\max_{1\leq i\leq n}\left\|\sqrt{T}b_{2i}\right\|^2$, it can be evaluated by considering the orders of magnitude of $\max_{1\leq i\leq n}\left\|\sqrt{T}\left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\left(\hat{F}-F\right)'M_{Xi}\epsilon_i\right\|^2$ and of $\max_{1\leq i\leq n}\left\|\sqrt{T}\left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma_i\right\|^2$. The former can be shown to be $o_p\left(n^{2/k_1}T\delta_{nT}^{-4}\right)$, based on the proof of Lemma A.6*(iii)*. The latter has the same order of magnitude as $\left\|T^{-1/2}\hat{F}'\left(\hat{F}-F\right)\right\|^2$, which is $O_p\left(T\delta_{nT}^{-4}\right)$ by Lemma A.3*(iii)*. Putting all together, $\max_{1\leq i\leq n} II_i = o_p\left(T^{3/2}n^{4/k_1}\delta_{nT}^{-6}\right)$ - so, $\max_{1\leq i\leq n} II_i$ is dominated by $\max_{1\leq i\leq n} I_i$. Similar passages yield that $\max_{1\leq i\leq n} III_i$ is dominated by $\max_{1\leq i\leq n} II_i$. Turning to $IV_i$, it holds that $\max_{1\leq i\leq n} IV_i \leq \left\|\sqrt{T}\left(\hat{\bar{\gamma}}-\bar{\gamma}\right)\right\|^2\left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|$, which is $o_p\left(Tn^{2/k_1}\delta_{nT}^{-6}\right)$ by Lemmas A.4 and A.7*(iv)*. Finally, $\max_{1\leq i\leq n} V_i$ is bounded by $\left\|\sqrt{T}\left(\hat{\bar{\gamma}}-\bar{\gamma}\right)\right\| \max_{1\leq i\leq n}\left\|\sqrt{T}b_{1i}\right\| \left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\| = o_p\left(Tn^{2/k_1}\delta_{nT}^{-4}\ln n\right)$. Putting all together, and using (52), it holds that

$$\begin{aligned}\max_{1\leq i\leq n} T\left(\hat{\gamma}_i-\hat{\bar{\gamma}}\right)'\hat{\Sigma}_{\gamma i}^{-1}\left(\hat{\gamma}_i-\hat{\bar{\gamma}}\right) &= \max_{1\leq i\leq n} N_i'\Sigma_{fMe,i}^{-1}N_i + o_p\left[(nT)^{1/k_1}\sqrt{\frac{\ln n}{T}}\right] \\ &\quad + o_p\left(\frac{n^{2/k_1}}{\sqrt{T}}\ln n\right) + o_p\left(\frac{\sqrt{T}n^{2/k_1}}{n}\ln n\right) + o_p(1),\end{aligned} \tag{53}$$

where the remainders are negligible as $(n,T)\to\infty$ with $\frac{(nT)^{1/k_1}}{\sqrt{T}} + \frac{\sqrt{T}n^{2/k_1}}{n} \to 0$ and $\frac{n^{4/k_1}}{T} \to 0$, which hold in light of (19). Finally, consider the sequence $\{N_i\}_{i=1}^n$: the covariance between $\sqrt{T}b_{1i}$ and $\sqrt{T}b_{1j}$ is given by

$$\begin{aligned} E\left(\frac{F'M_{Xi}\epsilon_i\epsilon_j'M_{Xj}F}{T}\right) &\leq E\left(\frac{F'\epsilon_i\epsilon_j'F}{T}\right) = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T E\left(f_tf_s'\epsilon_{it}\epsilon_{js}\right) \\ &\leq \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T \left\|E\left(f_tf_s'\right)\right\|\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| \leq M\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right|,\end{aligned}$$

which tends to zero as $(n,T)\to\infty$ by Assumption 7. By virtue of the asymptotic independence between $N_i$ and $N_j$ for all $i\neq j$, the asymptotics of $\max_{1\leq i\leq n} N_i'\Sigma_{fMe,i}^{-1}N_i$ is studied e.g. in Embrechts, Klüppelberg and Mikosch (1997, Table 3.4.4, p.156). Thus, equation (20) follows from (53).

We now finish the proof of the Theorem, analysing the power properties of the test. In order to evaluate the presence of power when $\gamma_i \neq \bar{\gamma}$ for some (at least one) $i$, after some algebra it can be shown that, under the alternative, $S_{\gamma,nT}$ has non-centrality parameter given by

$$S_{\gamma,nT}^{NC} = T \max_{1 \leq i \leq n} c_i' \hat{\Sigma}_{\gamma i}^{-1} c_i + 2T \max_{1 \leq i \leq n} c_i' \hat{\Sigma}_{\gamma i}^{-1} \left(\hat{\gamma}_i - \gamma_i\right) - 2T \max_{1 \leq i \leq n} c_i' \hat{\Sigma}_{\gamma i}^{-1} \left(\widehat{\bar{\gamma}} - \bar{\gamma}\right) = I + II - III,$$

with $I = O_p\left(T \left\|c_i\right\|^2\right)$ by construction. Also, $II$ is bounded by $\sqrt{T}\left(\max_{1 \leq i \leq n} \left\|c_i\right\|\right)\left(\max_{1 \leq i \leq n} \sqrt{T} \left\|\hat{\gamma}_i - \gamma_i\right\|\right) = O_p\left[T \delta_{nT}^{-2} n^{1/k_1} \left\|c_i\right\|\right]$ in view of Lemma A.6*(iii)*; similarly, $III = O_p\left(\sqrt{T} \delta_{nT}^{-2} \left\|c_i\right\|\right)$ by Lemma A.4. Let $S_{nT}^{\gamma,0}$ denote the null distribution of $S_{nT}^{\gamma}$; under $H_1^a$ it holds that

$$P\left[S_{\gamma,nT} > c_{\alpha,n}\right] = P\left[S_{nT}^{\gamma,0} > c_{\alpha,n} - S_{nT}^{\gamma,NC}\right],$$

which tends to 1 if $c_{\alpha,n} - S_{\gamma,nT}^{NC} \to -\infty$ as $(n,T) \to \infty$. In view of equation (22), we know that $c_{\alpha,n} = O(\ln n)$, whence (21) follows. QED

**Proof of Theorem 4**. The proof is very similar, in spirit, to the proof of Theorem 3, and therefore some passages are omitted to save space. Consider the following preliminary notation and derivations. We write

$$\hat{f}_t - \widehat{\bar{f}} = \left(f_t - f\right) + \left(\hat{f}_t - f_t\right) - \left(\widehat{\bar{f}} - f\right) = a_t + b_t - c_t.$$

Under $H_0^b$, $a_t = 0$ and $b_t = \hat{f}_t - f$; using (47), we can write

$$b_t = \left(\frac{\hat{F}'F}{T}\right) \frac{1}{n} \sum_{i=1}^{n} \gamma_i \epsilon_{it} + b_{2t} = b_{1t} + b_{2t}, \tag{54}$$

where $b_{2t}$ contains terms $I$–$VI$ and $VIII$ in (47). Also, for each $t$, $\hat{\Sigma}_{ft}^{-1} = \Sigma_{ft}^{-1} - \Sigma_{ft}^{-1}\left(\hat{\Sigma}_{ft} - \Sigma_{ft}\right)\Sigma_{ft}^{-1}$ $+ o_p\left(\left\|\hat{\Sigma}_{ft} - \Sigma_{ft}\right\|\right)$.

Neglecting higher order terms containing $o_p\left(\left\|\hat{\Sigma}_{ft} - \Sigma_{ft}\right\|\right)$, we have

$$\begin{aligned}
& n\left(\hat{f}_t - \widehat{\bar{f}}\right)' \hat{\Sigma}_{ft}^{-1} \left(\hat{f}_t - \widehat{\bar{f}}\right) \qquad (55) \\
= \; & n\left(b_{1t}' \Sigma_{ft}^{-1} b_{1t}\right) + n b_{1t}' \Sigma_{ft}^{-1} \left(\hat{\Sigma}_{ft} - \Sigma_{ft}^{-1}\right) \Sigma_{ft}^{-1} b_{1t} + n b_{2t}' \Sigma_{ft}^{-1} b_{2t} \\
& + 2n b_{1t}' \Sigma_{ft}^{-1} b_{2t} + n\left(\widehat{\bar{f}} - f\right)' \hat{\Sigma}_{ft}^{-1} \left(\widehat{\bar{f}} - f\right) - 2n\left(\widehat{\bar{f}} - f\right)' \hat{\Sigma}_{ft}^{-1} \left(\hat{f}_t - f\right) \\
= \; & n\left(b_{1t}' \Sigma_{ft}^{-1} b_{1t}\right) + I_t + II_t + III_t + IV_t - V_t.
\end{aligned}$$

After this preliminary calculations, we now turn to proving (24). Similarly to the proof of Theorem 3, we firstly prove that $\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)$ can be approximated by the maximum of a sequence of random variables with a $\chi_r^2$ distribution, up to a negligible error. Secondly, we show that, in (55), $\max_{1\leq t\leq T} I_t$, ..., $\max_{1\leq t\leq T} V_t$ are all $o_p(\ln T)$ uniformly in $t$.

We start from $\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)$. We show that the sequence $\{\sqrt{n}b_{1t}\}_{t=1}^{T}$ can be approximated by a sequence of *i.i.d.* Gaussian random variables with covariance matrix $\Sigma_{ft}$. To show this, recall that by Lemma A.7*(vi)*, we can write $n^{-1/2}\sum_{i=1}^{n}\gamma_i\epsilon_{it} = N_t + R_{Nt}$, with $N_t$ defined in Lemma A.7 as being zero mean Gaussian with covariance matrix $\Sigma_{\Gamma\epsilon,t}$, and $R_{Nt} = o_p\left(T^{1/k_2}n^{1/2-1/k_2}\right) + o_p\left(T^{1/k_2}\delta_{nT}^{-1}\sqrt{\ln T}\right) + o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\sqrt{\ln T}\right)$. Further, $T^{-1}\hat{F}'F = \Sigma_f + \left(T^{-1}F'F - \Sigma_f\right) + T^{-1}\left(\hat{F} - F\right)'F = \Sigma_f + R_f$, with $R_f = O_p\left(T^{-1/2}\right) + O_p\left(n^{-1}\right)$ by Lemmas A.7*(i)* and A.3*(ii)*. Hence

$$\sqrt{n}b_{1t} = \left(\Sigma_f + R_f\right)\left(N_t + R_{Nt}\right), \tag{56}$$

and

$$\begin{aligned} n\left(b_{1t}'\hat{\Sigma}_{ft}^{-1}b_{1t}\right) &= N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t + 2N_t'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_{Nt} + 2R_{Nt}'\Sigma_{\Gamma\epsilon,t}^{-1}N_t \\ &\quad +2N_t'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_fN_t + 2N_t'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_fR_{Nt} \\ &\quad +R_{Nt}'\Sigma_{\Gamma\epsilon,t}^{-1}R_{Nt} + 2R_{Nt}'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_fR_{Nt} \\ &\quad +N_t'R_f\Sigma_{ft}^{-1}R_fN_t + 2N_t'R_f\Sigma_{ft}^{-1}R_fR_{Nt} \\ &\quad +R_{Nt}R_f\Sigma_{ft}^{-1}R_fR_{Nt} \\ &= N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t + I_t^{b1} + II_t^{b1} + III_t^{b1} + IV_t^{b1} + V_t^{b1} + VI_t^{b1} \\ &\quad +VII_t^{b1} + VIII_t^{b1} + IX_t^{b1}. \end{aligned} \tag{57}$$

Passages are very similar to those after (51) in the proof of Theorem 3. In particular, it can be shown using Lemma A.7 that: $\max_{1\leq t\leq T} I_t^{b1}$ and $\max_{1\leq t\leq T} II_t^{b1}$ are both $o_p\left(T^{1/k_2}n^{1/2-1/k_2}\sqrt{\ln T}\right) + o_p\left(T^{1/k_2}\delta_{nT}^{-1}\sqrt{\ln T}\right) + o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\sqrt{\ln T}\right)$; $\max_{1\leq t\leq T} III_t^{b1} = O_p\left(T^{-1/2}\ln T\right) + O_p\left(n^{-1}\ln T\right)$; and that $\max_{1\leq t\leq T} IV_t^{b1}$,..., $\max_{1\leq t\leq T} IX_t^{b1}$ are all dominated and therefore negligible. Thus

$$\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right) = \max_{1\leq t\leq T} N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t + o_p\left[\sqrt{\frac{\ln T}{n}}\left(nT\right)^{1/k_2}\right] + o_p\left(T^{1/k_2}\frac{\sqrt{n\ln T}}{T}\right) + o_p(1), \tag{58}$$

where the approximation errors are negligible as long as $(n, T) \to \infty$ with $\frac{T^{4/k_2}}{n} \to 0$ and $T^{1/k_2}\frac{\sqrt{n}}{T} \to 0$.

After showing that $\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)$ can be approximated by $\max_{1\leq t\leq T} N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t$, we turn back to equation (55). We show that $\max_{1\leq t\leq T} I_t$, ..., $\max_{1\leq t\leq T} V_t$ in (55) are all $o_p(\ln T)$. We have that $\max_{1\leq t\leq T} I_t \leq \max_{1\leq t\leq T}\left\|\sqrt{n}b_{1t}\right\|^2 \max_{1\leq t\leq T}\left\|\Sigma_{ft}^{-1}\left(\hat{\Sigma}_{ft}-\Sigma_{ft}^{-1}\right)\Sigma_{ft}^{-1}\right\| = o_p\left(T^{2/k_2}\delta_{nT}^{-1}\ln T\right)$ by using Lemma A.7*(iii)*. Also, combining Lemmas A.5 and Lemma A.7*(iii)*, we have $\max_{1\leq t\leq T} IV_t = O_p\left(n\delta_{nT}^{-4}\right)+o_p\left(nT^{2/k_2}\delta_{nT}^{-5}\right)$ and $\max_{1\leq t\leq T} V_t = O_p\left(n\delta_{nT}^{-3}\right)+o_p\left(nT^{2/k_2}\delta_{nT}^{-4}\right)$. As far as $\max_{1\leq t\leq T} II_t$ and $\max_{1\leq t\leq T} III_t$ are concerned, studying their order of magnitude involves finding a bound for $\max_{1\leq t\leq T}\|b_{2t}\|$ and $\max_{1\leq t\leq T}\|b_{2t}\|^2$. Recall that

$$
\begin{aligned}
b_{2t} \;=\;& \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'x_{jt}-\frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'x_{jt}\\
&-\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j-\beta_j\right)'x_{jt}-\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'f_t\\
&-\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\epsilon_{jt}+\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\gamma_j'f_t+\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\epsilon_{jt}.
\end{aligned}
$$

Similar passages as in the proof of Lemma A.6*(ii)* yield $\max_{1\leq t\leq T}\|b_{2t}\| = o_p\left(T^{1/k_2}\delta_{nT}^{-2}\right)$ and $\max_{1\leq t\leq T}\|b_{2t}\|^2 = o_p\left(T^{2/k_2}\delta_{nT}^{-4}\right)$. We now turn to analyzing $\max_{1\leq t\leq T} II_t$ and $\max_{1\leq t\leq T} III_t$. As far as the former is concerned, $\max_{1\leq t\leq T} II_t \leq n\max_{1\leq t\leq T}\|b_{2t}\|^2 = o_p\left(T^{2/k_2}\phi_{nT}^{-2}\right)$. Also, $\max_{1\leq t\leq T} III_t \leq \sqrt{n}$ $\max_{1\leq t\leq T}\left\|\sqrt{n}b_{1t}\right\|\max_{1\leq t\leq T}\|b_{2t}\| = \sqrt{n}\ o_p\left(T^{1/k_2}\delta_{nT}^{-2}\sqrt{\ln T}\right)$. Putting all together, we have

$$
\begin{aligned}
\max_{1\leq t\leq T} n\left(\hat{f}_t-\widehat{\tilde{f}}\right)'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t-\widehat{\tilde{f}}\right) \;=\;& \max_{1\leq t\leq T} N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t+o_p\left[\sqrt{\frac{\ln T}{n}}\,(nT)^{1/k_2}\right]\\
&+o_p\left(\frac{T^{2/k_2}}{\sqrt{n}}\right)+o_p\left(T^{1/k_2}\frac{\sqrt{n\ln T}}{T}\right)+o_p(1);
\end{aligned}
\tag{59}
$$

under (23), the error term is negligible. Consider the sequence $\{N_t\}_{t=1}^T$. The covariance between $N_t$ and $N_{t-k}$ is proportional to, for $(n,T)\to\infty$

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(\gamma_i\gamma_j'\epsilon_{it}\epsilon_{jt-k}\right)\\
\leq\;& \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(\gamma_i\gamma_j'\epsilon_{it}\epsilon_{jt-k}\right)\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\gamma_i\gamma_j'\right\|\left|E\left(\epsilon_{it}\epsilon_{jt-k}\right)\right|\\
\leq\;& M\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|E\left(\epsilon_{it}\epsilon_{jt-k}\right)\right|,
\end{aligned}
$$

so that, under Assumption 9, $\lim_{k,n\to\infty} E\left(N_tN_{t-k}\right)\ln k = 0$. By virtue of such Berman condition, equation (24) holds - see e.g. Theorem 3.5.1 in Leadbetter and Rootzen (1988, p.470).

We now complete the proof of the Theorem by studying the power versus local alternatives. Under $H_1^b$, it can be shown that $S_{f,nT}$ has non-centrality parameter given by

$$\begin{aligned} S_{f,nT}^{NC} &= n \max_{1\leq t\leq T} c_t' \hat{\Sigma}_{ft}^{-1} c_t + 2n \max_{1\leq t\leq T} c_t' \hat{\Sigma}_{ft}^{-1} \left(\hat{f}_t - f_t\right) - 2n \max_{1\leq t\leq T} c_t' \hat{\Sigma}_{ft}^{-1} \left(\widehat{\overline{f}} - f\right) \\ &= I + II + III, \end{aligned}$$

with $I = O_p\left(n \left\|c_t\right\|^2\right)$ by construction. Also, $II$ is bounded by $n \left(\max_{1\leq t\leq T} \left\|c_t\right\|\right) \max_{1\leq t\leq T} \left\|\left(\hat{f}_t - f_t\right)\right\| = O_p\left(n \left\|c_t\right\| T^{2/k_2} \delta_{nT}^{-2}\right)$ by Lemma A.6*(ii)*; similarly, $III = O_p\left(n\delta_{nT}^{-2} \left\|c_t\right\|\right)$. Let $S_{nT}^{f,0}$ denote the null distribution of $S_{f,nT}$. Then, under $H_1^b$ we have

$$P\left[S_{f,nT} > c_{\alpha,T}\right] = P\left[S_{nT}^{f,0} > c_{\alpha,T} - S_{f,nT}^{NC}\right];$$

$P\left[S_{f,nT} > c_{\alpha,T}\right]$ tends to 1 if $c_{\alpha,T} - S_{f,nT}^{NC} \to -\infty$ as $(n,T) \to \infty$; this holds because, by (26), $c_{\alpha,T} = O\left(\ln T\right)$. QED

**Proof of Theorem 5**. We start with $\tilde{S}_{\gamma,nT}$. Under $H_0^a$ we have

$$\begin{aligned} \sqrt{2r}\tilde{S}_{\gamma,nT} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[T\left(\hat{\gamma}_i - \gamma\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\hat{\gamma}_i - \gamma\right) - r\right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T\left(\widehat{\overline{\gamma}} - \gamma\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\widehat{\overline{\gamma}} - \gamma\right) \quad (60) \\ &\quad - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} T\left(\hat{\gamma}_i - \gamma\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\widehat{\overline{\gamma}} - \gamma\right) \\ &= I + II - III. \end{aligned}$$

Consider $I$; using (43), we can write

$$
\begin{aligned}
I &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\epsilon_i' M_{Xi}F}{\sqrt{T}}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{F'M_{Xi}\epsilon_i}{\sqrt{T}}-r\right] \\
&+\frac{1}{\sqrt{n}}T\sum_{i=1}^{n}\frac{\epsilon_i' M_{Xi}\left(\hat{F}-F\right)}{T}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{\left(\hat{F}-F\right)'M_{Xi}\epsilon_i}{T} \\
&+\frac{1}{\sqrt{n}}T\sum_{i=1}^{n}\frac{\gamma'\left(\hat{F}-F\right)'M_{Xi}\hat{F}}{T}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma}{T} \\
&+\frac{2}{\sqrt{n}}\sqrt{T}\sum_{i=1}^{n}\frac{\epsilon_i' M_{Xi}\left(\hat{F}-F\right)}{T}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{F'M_{Xi}\epsilon_i}{\sqrt{T}} \\
&+\frac{2}{\sqrt{n}}T\sum_{i=1}^{n}\frac{\epsilon_i' M_{Xi}\left(\hat{F}-F\right)}{T}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma}{T} \\
&+\frac{2}{\sqrt{n}}\sqrt{T}\sum_{i=1}^{n}\frac{\gamma'\left(\hat{F}-F\right)'M_{Xi}\hat{F}}{T}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{F'M_{Xi}\epsilon_i}{\sqrt{T}} \\
&= I_a+I_b+I_c+I_d+I_e+I_f.
\end{aligned}
$$

By (29), $I_a$ is $O_p(1)$. Turning to $I_b$, it is bounded by

$$
\sqrt{n}TE\left[\frac{\epsilon_i' M_{Xi}\left(\hat{F}-F\right)}{T}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\hat{\Sigma}_{\gamma i}^{-1}\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{\left(\hat{F}-F\right)'M_{Xi}\epsilon_i}{T}\right]\leq M\sqrt{n}TE\left\|\frac{\epsilon_i'\left(\hat{F}-F\right)}{T}\right\|^2,
$$

where we have used the consistency of $\hat{\Sigma}_{\gamma i}$ and Assumptions 3*(i)* and 4*(i)*. Applying Lemma A.2*(i)*, we have $I_b = O_p\left(\sqrt{n}T\delta_{nT}^{-4}\right)$. By a similar logic, it can be shown that $I_c$ is bounded by $\sqrt{n}TE\left\|T^{-1}\left(\hat{F}-F\right)F'\right\|^2$, which is $O_p\left(\sqrt{n}T\delta_{nT}^{-4}\right)$ by virtue of Lemma A.3*(ii)*. Turning to $I_d$, similar passages as above entails that it is bounded by

$$
\begin{aligned}
\sqrt{n}TE\left[\frac{\epsilon_i' M_{Xi}\left(\hat{F}-F\right)}{T}\frac{F'M_{Xi}\epsilon_i}{\sqrt{T}}\right] &\leq \sqrt{nT}\left[E\left(\left\|\frac{\epsilon_i'\left(\hat{F}-F\right)}{T}\right\|^2\right)\right]^{1/2}\left[E\left(\left\|\frac{F'\epsilon_i}{\sqrt{T}}\right\|^2\right)\right]^{1/2} \\
&= \sqrt{nT}O_p\left(\delta_{nT}^{-2}\right)O_p(1).
\end{aligned}
$$

Similarly, $I_e$ is bounded by

$$
\begin{aligned}
& \sqrt{n}TE\left[\frac{\epsilon_i' M_{Xi}\left(\hat{F}-F\right)}{T}\frac{\hat{F}' M_{Xi}\left(\hat{F}-F\right)\gamma}{T}\right] \\
\leq \quad & \sqrt{n}T\left[E\left(\left\|\frac{\epsilon_i'\left(\hat{F}-F\right)}{T}\right\|^2\right)\right]^{1/2}\left[E\left(\left\|\frac{\hat{F}'\left(\hat{F}-F\right)}{T}\right\|^2\right)\right]^{1/2} = \sqrt{n}TO_p\left(\delta_{nT}^{-2}\right)O_p\left(\delta_{nT}^{-2}\right);
\end{aligned}
$$

using a similar logic, it can be shown that $I_f = O_p\left(\sqrt{nT}\delta_{nT}^{-2}\right)$. Putting all together, $I = O_p(1)$ $+O_p\left(\sqrt{nT}\delta_{nT}^{-2}\right)$ $+O_p\left(\sqrt{n}T\delta_{nT}^{-4}\right)$. Finally, consider $II$ and $III$ in (60). As far as $II$ is concerned, note that $II = \sqrt{n}T\left(\widehat{\bar{\gamma}}-\bar{\gamma}\right)'\Sigma_{\gamma i}^{-1}\left(\widehat{\bar{\gamma}}-\bar{\gamma}\right)$ $+o_p(1)$ by consistency of $\hat{\Sigma}_{\gamma i}$. Thus, Lemma A.4 entails that $II = O_p\left(\sqrt{n}T\delta_{nT}^{-4}\right)$. Turning to $III$, this is bounded by $\sqrt{n}T\max_i\Sigma_{\gamma i}^{-1}\left\|\widehat{\bar{\gamma}}-\bar{\gamma}\right\|\left\|\frac{1}{n}\sum_{i=1}^n\left(\hat{\gamma}_i-\bar{\gamma}\right)\right\|$, which has the same order of magnitude as $II$. Putting all together, it holds that $\sqrt{2r}\tilde{S}_{\gamma,nT} = O_p(1)$ $+O_p\left(\sqrt{nT}\delta_{nT}^{-2}\right)$ $+O_p\left(\sqrt{n}T\delta_{nT}^{-4}\right) = O_p(1)$ $+O_p\left(\sqrt{\frac{n}{T}}\right)$ $+O_p\left(\sqrt{\frac{T}{n}}\right)$.

We now turn to studying $\tilde{S}_{f,nT}$. Under $H_0^b$, we have

$$
\begin{aligned}
\sqrt{2r}\tilde{S}_{f,nT} &= \frac{1}{\sqrt{T}}\sum_{t=1}^T\left[n\left(\hat{f}_t-f\right)'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t-f\right)-r\right] \\
&\quad +\frac{1}{\sqrt{T}}\sum_{t=1}^T n\left(\widehat{\bar{f}}-f\right)'\hat{\Sigma}_{ft}^{-1}\left(\widehat{\bar{f}}-f\right)-\frac{2}{\sqrt{T}}\sum_{t=1}^T n\left(\hat{f}_t-f\right)'\hat{\Sigma}_{ft}^{-1}\left(\widehat{\bar{f}}-f\right) \\
&= I+II-III.
\end{aligned} \tag{61}
$$

Consider $I$; using (54) we may write

$$
I = \frac{1}{\sqrt{T}}\sum_{t=1}^T\left(nb_{1t}'\hat{\Sigma}_{ft}^{-1}b_{1t}-r\right)+\frac{1}{\sqrt{T}}\sum_{t=1}^T nb_{2t}'\hat{\Sigma}_{ft}^{-1}b_{2t}+\frac{2}{\sqrt{T}}\sum_{t=1}^T nb_{1t}'\hat{\Sigma}_{ft}^{-1}b_{2t} = I_a+I_b+I_c.
$$

By (30), $T^{-1/2}\sum_{t=1}^T\left(nb_{1t}'\hat{\Sigma}_{ft}^{-1}b_{1t}-r\right) = O_p(1)$. As far as $I_b$ is concerned, by virtue of the consistency of $\Sigma_{ft}$, it is bounded by $n\sqrt{T}E\left\|b_{2t}\right\|^2 = n\sqrt{T}\min\left\{T^{-1},n^{-2}\right\} = O_p\left(\frac{n}{\sqrt{T}}\right)+O_p\left(\frac{\sqrt{T}}{n}\right)$, which follows from the proof of Theorem 2. Finally, turning to $I_c$ and setting $\hat{\Sigma}_{Ft}^{-1} = I_r$ for simplicity, we may write

$$
\begin{aligned}
\frac{1}{2}I_c &= n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{jt}\epsilon_{it}\right) \\
&\quad -n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\left(\frac{\hat{F}'F}{T}\right)\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{jt}\epsilon_{it}\right) \\
&\quad -n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\frac{1}{T}\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{jt}\epsilon_{it}\right) \\
&\quad -n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\epsilon_{it}\right) \\
&\quad -n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt}\epsilon_{it}\right) \\
&\quad +n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\frac{\hat{F}'\epsilon_j}{T}\gamma_j'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\epsilon_{it}\right) \\
&\quad +n\left(\frac{F'\hat{F}}{T}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i'\frac{\hat{F}'\epsilon_j}{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt}\epsilon_{it}\right) \\
&= I_{c,1}-I_{c,2}-I_{c,3}-I_{c,4}-I_{c,5}+I_{c,6}+I_{c,7}.
\end{aligned}
$$

Studying the order of magnitude of each term is based on similar passages to the ones in the proof of Theorem 2. The only differences are: the summation across $t$; the normalization by $T^{-1/2}$; and the multiplication by $n$. The effect of summing across $t$ is washed out by the normalization by $T^{-1/2}$ for all terms $I_{c,1}-I_{c,4}$ and $I_{c,6}$, which can be shown by the same arguments as in (39). We have $I_{c,1}=O_p\left(\sqrt{n}T^{-1}\right)$; $I_{c,2}=O_p\left(\delta_{nT}^{-1}\right)$; $I_{c,3}=O_p\left(\sqrt{n}T^{-1}\right)$, $I_{c,4}=O_p\left(\delta_{nT}^{-1}\right)$ and $I_{c,6}=O_p\left(T^{-1/2}\right)$. As far as $I_{c,5}$ and $I_{c,7}$ are concerned, the contribution of $T^{-1/2}\sum_{t=1}^{T}\epsilon_{jt}\epsilon_{it}$ is at most $O_p\left(\sqrt{T}\right)$; thus, $I_{c,5}=O_p\left(\sqrt{n}\right)$ and $I_{c,7}=O_p\left(\sqrt{T}\delta_{nT}^{-2}\right)$. We now turn to analyzing $II$ and $III$ in (61). By Lemma A.5, $II=n\sqrt{T}O_p\left(\delta_{nT}^{-4}\right)$. As far as $III$ is concerned, using the consistency of $\hat{\Sigma}_{ft}$ and the invertibility of $\Sigma_{ft}$, it is bounded by $n\sqrt{T}\max_t\left\|\Sigma_{ft}^{-1}\right\|\left\|\widehat{f}-f\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{f}_t-f\right)\right\|=n\sqrt{T}O_p\left(\delta_{nT}^{-4}\right)$, again by Lemma A.5. Putting all together, the result follows. QED

**Proof of Theorem 6**. We report the proof for $S_{\gamma,nT}^{CCE}$ only - the proof for $S_{\gamma,nT}^{IE}$ is almost identical; the only difference is the need for the restriction $\frac{\sqrt{n}}{T}\to 0$, which can be shown based on the passages in the proof of Theorem 7.

Consider the building block of the test statistic, viz. $\hat{\beta}^{bw}-\beta$:

$$\begin{aligned}\hat{\beta}^{bw}-\beta &= \frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}x_{it}'(\beta_i-\beta)\right]\\ &-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jt}\dot{x}_{jt}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jt}x_{it}'(\beta_i-\beta)\right]\\ &+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\epsilon_{it}\right].\end{aligned}$$

Note also that

$$\begin{aligned}\hat{\beta}^{CCE}-\beta &= \frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_w\epsilon_i}{T}\right)\\ &+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_wF}{T}\gamma_i\right),\end{aligned}$$

so that

$$\begin{aligned}\hat{\beta}^{bw}-\hat{\beta}^{CCE} &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}x_{it}'(\beta_i-\beta)\right]\\ &-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jt}\dot{x}_{jt}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jt}x_{it}'(\beta_i-\beta)\right]\\ &+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\epsilon_{it}\right]-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_w\epsilon_i}{T}\right)\\ &-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_wF}{T}\gamma_i\right)\\ &= I+II+III-IV-V. \qquad (62)\end{aligned}$$

Terms $IV+V$ have magnitude $O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{n}\right)$, as discussed above. Also, in a similar way it can be shown that $III=O_p\left(\frac{1}{\sqrt{nT}}\right)$. Finally, we have

$$I+II=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{it}x_{it}'(\beta_i-\beta),$$

where $\tilde{D}_{it}=\left[T^{-1}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\dot{x}_{it}-n^{-1}\sum_{j=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jt}\dot{x}_{jt}'\right]^{-1}\dot{x}_{jt}$. By Assumption 3, the sequence $T^{-1}\sum_{t=1}^{T}\tilde{D}_{it}x_{it}'(\beta_i-\beta)$ is uncorrelated across $i$, so that the magnitude of $I+II$ is proportional to the square root of $n^{-2}\sum_{i=1}^{n}E\left\|\tilde{D}_{it}x_{it}'(\beta_i-\beta)\right\|^2\le n^{-2}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}E\left\|\tilde{D}_{it}x_{it}'\right\|^2$ $E\left\|\beta_i-\beta\right\|^2$. Using Assumptions 3 and 2*(i)*, this is of order $O\left(n^{-1}\right)$,so that $I+II=O_p\left(n^{-1/2}\right)$. The limiting distribution follows from standard arguments, upon noting that the sequence $T^{-1}\sum_{t=1}^{T}\tilde{D}_{it}x_{it}'(\beta_i-\beta)$ is conditionally independent across $i$ by Assumption 3, and has finite moment of order $2+\delta$ for $\delta>0$. Putting all together, the null distribution follows.

As far as power is concerned, the CCE estimator is consistent under alternatives; as far as the between estimator is concerned, $\hat{\beta}^{bw}-\beta$ has the same expansion as above with the extra term

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}f_t'\left(\gamma_i-\frac{1}{n}\sum_{i=1}^{n}\gamma_i\right)\right] \\
= & \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}f_t'(\gamma_i-\gamma)\right]-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}f_t'\left(\frac{1}{n}\sum_{i=1}^{n}\gamma_i-\gamma\right)\right] \\
= & I+II,
\end{aligned}
$$

where term $II$ is clearly dominated. As far as $I$ is concerned, when premultiplied by $\sqrt{n}$, we have $I$ $= n^{-1/2}\sum_{i=1}^{n}C_{iT}$. By assumption, $C_{iT}$ has mean zero, it can be shown to have finite moment of order $2+\delta$ for $\delta>0$, and it is conditionally independent across $i$. It is also conditionally independent of $II+III$ in (62). This entails that, under alternatives, $\sqrt{n}\left(\hat{\beta}^{bw}-\hat{\beta}^{CCE}\right)$ converges to a normally distributed random variable with mean zero, and a higher variance than under the null. Standard passages ensure the validity of the theorem. QED

**Proof of Theorem 7**. Consider first equation (32); we start with $\sqrt{nT}\left(\hat{\beta}^{CCE}-\hat{\beta}^{FE}\right)$ under $H_0^a$. Recall that under the null $H_0^b$, $f_t=f=ci_T$, where $c$ is a constant. Therefore, $M_F=I_T-T^{-1}i_Ti_T'$. This entails that

$$
\hat{\beta}^{FE}-\beta=\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'M_FX_i}{T}\right)^{-1}\left(\frac{X_i'M_F\epsilon_i}{T}\right). \tag{63}
$$

By using (38) and equation (56) in Pesaran (2006, p. 982):

$$
\begin{aligned}
\sqrt{nT}\left(\hat{\beta}^{CCE}-\hat{\beta}^{FE}\right) = & \sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)+\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_w\epsilon_i}{T}\right) \\
& +\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_wF}{T}\gamma_i\right) \\
& -\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\beta)-\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'M_FX_i}{T}\right)^{-1}\left(\frac{X_i'M_F\epsilon_i}{T}\right).
\end{aligned}
$$

Under the rank condition in Assumption 4*(ii)*, we have that

$$\begin{aligned}\frac{X_i'\bar{M}_w\epsilon_i}{T} &= \frac{X_i'M_F\epsilon_i}{T}+O_p\left(\frac{1}{n}\right),\\ \frac{X_i'\bar{M}_wF}{T} &= O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right);\end{aligned}$$

note that the former equation does not need the rank condition in Assumption 4*(ii)*, whereas the latter does. Therefore

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}^{CCE}-\hat{\beta}^{FE}\right) &= O_p\left(\sqrt{\frac{T}{n}}\right)+\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\left(\frac{X_i'\bar{M}_wF}{T}\gamma_i\right)\\ &= O_p\left(\sqrt{\frac{T}{n}}\right)+O_p\left(\sqrt{\frac{T}{n}}\right)+O_p\left(1\right),\end{aligned}$$

which proves part 1 of the Theorem. The asymptotics of the terms $O_p\left(\sqrt{\frac{T}{n}}\right)$ and $O_p\left(1\right)$ depends on the DGP of $x_{it}$ and $y_{it}$ through $\bar{M}_w$ and $T^{-1}\left(X_i'\bar{M}_wF\right)$.

Consider now equation (31). Note that

$$\hat{\beta}^{IE}-\beta=\left(\hat{\beta}^{IE}-\frac{1}{n}\sum_{i=1}^{n}\beta_i\right)+\left(\frac{1}{n}\sum_{i=1}^{n}\beta_i-\beta\right),$$

so that, using (63):

$$\begin{aligned}\sqrt{nT}\left(\hat{\beta}^{IE}-\hat{\beta}^{FE}\right) &= \sqrt{nT}\left(\hat{\beta}^{IE}-\frac{1}{n}\sum_{i=1}^{n}\beta_i\right)-\sqrt{nT}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i'M_FX_i}{T}\right)^{-1}\left(\frac{X_i'M_F\epsilon_i}{T}\right)\\ &= \sqrt{nT}\left(\hat{\beta}^{IE}-\frac{1}{n}\sum_{i=1}^{n}\beta_i\right)+O_p\left(1\right),\end{aligned}$$

where the $O_p\left(1\right)$ term holds by Assumptions 1 and 2. Let $\Gamma=[\gamma_1|...|\gamma_n]$; using equation (42) in Song

(2013), we have

$$
\begin{aligned}
& \hat{\beta}^{IE} - \frac{1}{n}\sum_{i=1}^{n}\beta_i \\
= \; & \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\beta}_i^{IE} - \beta_i\right) \\
= \; & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1}\left(\frac{X_i' M_{\hat{F}}\epsilon_i}{T}\right) \\
& -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_j}{T}\right)\left[\gamma_j'\left(\frac{\Gamma'\Gamma}{n}\right)^{-1}\gamma_i\right]\left(\tilde{\beta}_j - \beta_j\right) \\
& -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}} X_j}{T}\left(\tilde{\beta}_j^{IE} - \beta_j\right)\left(\tilde{\beta}_j^{IE} - \beta_j\right)'\frac{X_j'\hat{F}}{T}\right. \\
& +\frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}} X_j}{T}\left(\tilde{\beta}_j^{IE} - \beta_j\right)\frac{\epsilon_j'\hat{F}}{T} + \frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}} F}{T}\gamma_j\left(\tilde{\beta}_j^{IE} - \beta_j\right)'\frac{X_j'\hat{F}}{T} + \\
& \frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}}\epsilon_j}{T}\left(\tilde{\beta}_j^{IE} - \beta_j\right)'\frac{X_j'\hat{F}}{T} + \frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}} F}{T}\gamma_j\frac{\epsilon_j'\hat{F}}{T} + \frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}}\epsilon_j}{T}\gamma_j'\frac{F'\hat{F}}{T} \\
& \left. +\frac{1}{n}\sum_{j=1}^{n}\frac{X_i' M_{\hat{F}}\epsilon_j}{T}\frac{\epsilon_j'\hat{F}}{T}\right]\left(\frac{F'\hat{F}}{T}\right)^{-1}\left(\frac{\Gamma'\Gamma}{n}\right)\gamma_i \\
= \; & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1}\left(\frac{X_i' M_{\hat{F}}\epsilon_i}{T}\right) + J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + J_7 + J_8;
\end{aligned}
$$

quantities like $\frac{F'\hat{F}}{T}$, $\frac{\Gamma'\Gamma}{n}$ and $\gamma_i$ will be omitted henceforth, to simplify the notation. Similar passages as above yield $\sqrt{nT}\ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1}\left(\frac{X_i' M_{\hat{F}}\epsilon_i}{T}\right) = O_p(1)$. As far as the other terms are concerned, note first that, by Proposition 1 in Song (2013), $\tilde{\beta}_i^{IE} - \beta_i = O_p\left(\phi_{nT}^{-1}\right)$. Since the order of magnitude of an average is bounded by the order of the summands, the same passages as in Song (2013) would entail $J_2 = O_p\left(\phi_{nT}^{-2}\right)$; $J_3$, $J_5$ and $J_6$ are all bounded by $O_p\left(\phi_{nT}^{-2}\right) + O_p\left(\phi_{nT}^{-1}\delta_{nT}^{-1}\right)$; $J_4 = O_p\left(\phi_{nT}^{-2}\right) + O_p\left(\phi_{nT}^{-1}\delta_{nT}^{-2}\right)$; $J_7 = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right) + O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Putting all together, this entails that $\sqrt{nT}\,(J_2 + J_3 + J_4 + J_5 + J_6 + J_7) = O_p\left(\sqrt{\frac{n}{T}}\right) + O_p\left(\sqrt{\frac{T}{n^3}}\right) + O_p(1)$. This bound is not the sharpest possible, but it suffices for our purposes. Finally, consider $J_1$ and $J_8$. As far as the former is concerned, we have

$$
J_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_j}{T}\right)\left[\gamma_j'\left(\frac{\Gamma'\Gamma}{n}\right)^{-1}\gamma_i\right]\left(\tilde{\beta}_j - \beta_j\right),
$$

which, by virtue of Assumptions 2*(i)* and 4*(i)*, has the same order as derived in Song (2013); thus,

$J_1 = O_p\left(n^{-1/2}T^{-1/2}\right)$ and $\sqrt{nT}J_2 = O_p\left(1\right)$. Turning to $J_8$

$$
\begin{aligned}
-J_8 &= \frac{1}{nT}\sum_{i=1}^{n}\left(\frac{X_i' M_{\hat{F}} X_i}{T}\right)^{-1} X_i' M_{\hat{F}}\left(\frac{1}{nT}\sum_{k=1}^{n}\epsilon_k\epsilon_k'\hat{F}\right) \\
&\leq M\frac{1}{nT^2}\sum_{k=1}^{n}\left\|X_i' M_{\hat{F}}\left(\epsilon_k\epsilon_k'\hat{F}\right)\right\| = O_p\left(\frac{1}{\delta_{nT}^2}\right) + O_p\left(\frac{1}{\phi_{nT}\sqrt{T}}\right),
\end{aligned}
$$

where the last passage comes from the proof of Proposition 1 in Song (2013). Therefore, $\sqrt{nT}J_8 = O_p\left(\sqrt{\frac{n}{T}}\right) + O_p\left(\sqrt{\frac{T}{n}}\right)$. Putting all together, part 2 of the Theorem follows. The behaviour of the test statistics under the null $H_0^a$ follows immediately from the passages above, since the term $\sqrt{nT}\left(\hat{\beta}^{IE} - \frac{1}{n}\sum_{i=1}^{n}\beta_i\right)$ is still $O_p\left(\sqrt{\frac{n}{T}}\right) + O_p\left(\sqrt{\frac{T}{n}}\right)$. QED

# References


Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica, vol. 59, 817-858.

Bai, J., 2003, Inferential theory for structural models of large dimensions. Econometrica, vol. 71, 135-171.

Bai, J., 2009a, Panel data models with interactive fixed effects. Econometrica, vol. 77, 1229-1279.

Bai, J., 2009b, Supplement to Panel data models with interactive fixed effects : technical details and proofs. Econometrica, vol. 77, 1-30.

Baltagi, B., Kao, C., Na, S., 2012, Testing cross-sectional dependence in panel factor model using the wild bootstrap F-test, manuscript.

Bai, J., Ng, S., 2002, Determining the number of factors in approximate factor models. Econometrica, vol. 70, 191-221.

Berman, S.M., 1964, Limit theorems for the maximum term in stationary sequences. The Annals of Mathematical Statistics , vol. 35, 502-516.

Berkes, I., Liu, W., Wu, W.B., 2013, Komlos-Major-Tusnady approximation under dependence. Annals of Probability (forthcoming).

Canto e Castro L., 1987, Uniform rate of convergence in extreme-value theory: normal and gamma models. Annales Scientifiques de l'Université de Clerment-Ferrand, 2, tome 90, Série Probabilités et Applications, vol. 6, 25-41.

Castagnetti, C., Rossi, E., 2013, Euro corporate bond risk factors. Journal of Applied Econometrics, vol. 28, 372-391.

Chudik, A., and Pesaran, H., 2013, Common Correlated Effects Estimation of Heterogenous Dynamic Panel Data Models with Weakly Exogenous Regressors. CESifo Working Paper Series 4232.

Chudik, A., Pesaran, H., Tosetti E., 2011, Weak and strong cross-section dependence and estimation of large panels. Econometrics Journal, vol. 14, 45-90.

Corradi, V., 1999, Deciding between I(0) and I(1) via flil-based bounds. Econometric Theory, vol. 15, 643-63.

Csörgö, M., Hórvath, L., 1997, Limit theorems in change-point analysis. Wiley, Chichester.

Csörgö, M., Révész, P., 1975, A new method to prove Strassen-type laws of invariance principle. I. Probability Theory and Related Fields, vol. 31, 255-260.

Csörgö, M., Révész, P., 1975, A new method to prove Strassen-type laws of invariance principle. II. Probability Theory and Related Fields, vol. 31, 261-269.

Davidson J., 1994, Stochastic Limit Theory. Oxford University Press.

Eberhardt, M., Helmers, C., Strauss, H., 2013, Do spillovers matter when estimating private returns to R&D?. The Review of Economics and Statistics, vol. 95, 436-448.

Eberhardt, M., Teal, F., 2012, No mangos in the tundra: spatial heterogeneity in agricultural productivity analysis. Oxford Bulletin of Economics and Statistics (forthcoming).

Eberlein, E., 1986, On strong invariance principles under dependence assumptions. Annals of Probability, vol. 14, 260-270.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997, Modelling extremal events for insurance and finance. New York: Springer.

Everaert, G., Groote, T.D., 2012. Common correlated effects estimation of dynamic panels with cross-sectional dependence. Mimeo.

French, D., O'Hare, C., 2013, A dynamic factor approach to mortality modeling. Journal of Forecasting (forthcoming).

Hall, P., Miller, H., 2010, Bootstrap confidence intervals and hypothesis tests for extrema of parameters. Biometrika, vol. 97, 881-892.

Hannan, E.T., Kavalieris, L., 1986. Regression; autoregression models. Journal of Time Series Analysis, vol. 7, 27-49.

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under Near-Epoch Dependence. Journal of Econometrics, vol. 170, 178-190.

Kao, C., Trapani, L., Urga, G., 2012, Testing for Instability in Covariance Structures. Center for Policy Research Working Papers No. 131.

Kapetanios, G., 2003, Determining the poolability of individual series in panel datasets. University of London Queen Mary Economics Working Paper No. 499.

Kapetanios, G., Pesaran, M.H., 2007, Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. In Garry Phillips and Elias Tzavalis, (Eds.), The Refinenement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, Cambridge.

Komlós, J., Major, P., Tusnády, G., 1975, An approximation of partial sums of independent rv's and the sample df. I. Probability Theory and Related Fields, vol. 32, 111-131.

Komlós, J., Major, P., Tusnády, G., 1976, An approximation of partial sums of independent rv's and the sample df. II. Probability Theory and Related Fields, vol. 34, 33-58.

Leadbetter, M.R., Rootzen, H., 1988, Extremal theory for stochastic processes. Annals of Probability, vol. 16, 431-478.

Lee, R. D., Carter, L. R., 1992, Modeling and forecasting the time series of U.S. mortality. Journal of the American Statistical Association, vol. 87, 659-671.

Lin, Z., Bai, Z., 2010. Probability Inequalities. Berlin: Springer.

Ling, S., 2007. Testing for change points in time series models and limiting theorems for NED sequences. Annals of Statistics, vol. 35, 1213-1237.

Peligrad, M., Utev, S., Wu, W.B., 2007. A maximal $L_p$-inequality for stationary sequences and application. Proceedings of the American Statistical Association, vol. 135, 541-550.

Pesaran, M. H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, vol. 74, 967-1012.

Pesaran, M. H., Tosetti, E., 2011, Large panels with common factors and spatial correlation. Journal of Econometrics, vol. 161, 182-202.

Pesaran, M.H., Yamagata, T., 2008, Testing slope homogeneity in large panels. Journal of Econometrics, vol. 142, 50-93.

Phillips, P.C.B., Moon, H. R., 1999, Linear regression limit theory for nonstationary panel data. Econometrica, vol. 67, 1057-1112.

Phillips, P.C.B., Solo, V., 1992, Asymptotics for linear processes. Annals of Statistics, vol. 20, 971-1001.

Sarafidis, V., Yamagata, T., Robertson, D., 2009, A test of cross section dependence for a linear dynamic panel model with regressors. Journal of Econometrics, vol. 148, 149-461.

Shorack, G.R., Wellner, J.A., 1986, Empirical processes with applications to statistics. Wiley, New York.

Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Mimeo, January 2013.

Strang, G., 1988, Linear algebra and its applications. Third Edition, Harcourt, Orlando.

Westerlund, J., Hess, W., 2011, A new poolability test for cointegrated panels. Journal of Applied Econometrics, vol. 26, 56-88.

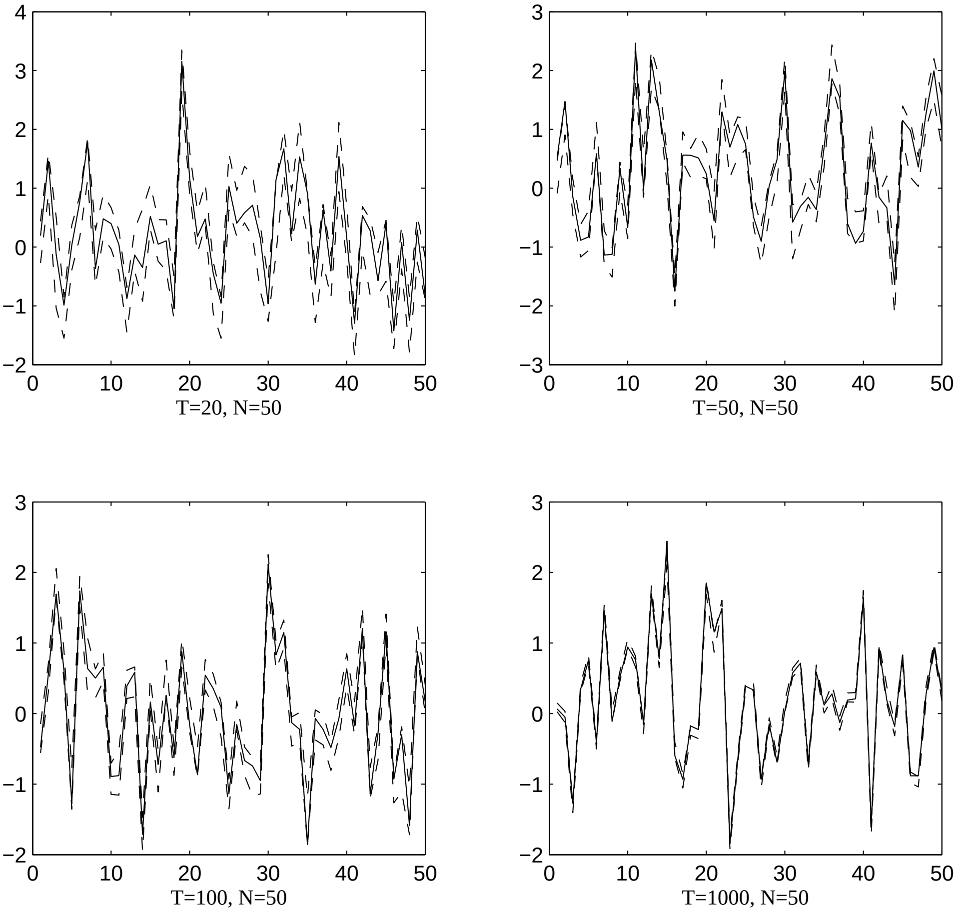


Figure 1: Confidence intervals for $\gamma_i$. For each value of $i = 1, ..., 50$ (on the horizontal axis), the solid line represents the true loading $\gamma_i$. The dashed lines are the confidence intervals at 95% confidence level for each $i$.

| $n$ \ $T$ | 30 | 50 | 100 | 200 |
|---|---|---|---|---|
| 30 | 0.977 | 0.964 | 0.979 | 0.974 |
| 50 | 0.976 | 0.963 | 0.989 | 0.970 |
| 100 | 0.991 | 0.987 | 0.988 | 0.991 |
| 200 | 0.992 | 0.994 | 0.997 | 0.997 |

Table 1: Average correlation coefficients between $\{\hat{f}_t\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$.

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.077 | 0.066 | 0.060 | 0.056 | 0.950 | 0.996 | 1.000 | 1.000 |
| 50 | 0.073 | 0.063 | 0.050 | 0.056 | 0.986 | 0.999 | 1.000 | 1.000 |
| 100 | 0.073 | 0.063 | 0.052 | 0.045 | 0.997 | 1.000 | 1.000 | 1.000 |
| 200 | 0.072 | 0.062 | 0.053 | 0.042 | 0.998 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.086 | 0.074 | 0.064 | 0.059 | 0.867 | 0.968 | 0.999 | 1.000 |
| 50 | 0.078 | 0.067 | 0.053 | 0.058 | 0.926 | 0.993 | 1.000 | 1.000 |
| 100 | 0.074 | 0.063 | 0.053 | 0.046 | 0.972 | 1.000 | 1.000 | 1.000 |
| 200 | 0.073 | 0.064 | 0.054 | 0.042 | 0.992 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.109 | 0.094 | 0.081 | 0.076 | 0.612 | 0.800 | 0.976 | 0.999 |
| 50 | 0.090 | 0.079 | 0.065 | 0.068 | 0.667 | 0.883 | 0.993 | 1.000 |
| 100 | 0.085 | 0.070 | 0.058 | 0.051 | 0.764 | 0.952 | 1.000 | 1.000 |
| 200 | 0.076 | 0.067 | 0.057 | 0.044 | 0.863 | 0.983 | 1.000 | 1.000 |

Table 2: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (33)- (34).

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.067 | 0.053 | 0.046 | 0.043 | 0.999 | 1.000 | 1.000 | 1.000 |
| 50 | 0.067 | 0.055 | 0.048 | 0.047 | 1.000 | 1.000 | 1.000 | 1.000 |
| 100 | 0.066 | 0.064 | 0.052 | 0.040 | 1.000 | 1.000 | 1.000 | 1.000 |
| 200 | 0.069 | 0.063 | 0.050 | 0.041 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.069 | 0.055 | 0.051 | 0.046 | 0.979 | 0.998 | 1.000 | 1.000 |
| 50 | 0.069 | 0.057 | 0.049 | 0.049 | 0.996 | 1.000 | 1.000 | 1.000 |
| 100 | 0.067 | 0.064 | 0.054 | 0.041 | 0.999 | 1.000 | 1.000 | 1.000 |
| 200 | 0.069 | 0.064 | 0.05 | 0.041 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.081 | 0.066 | 0.060 | 0.052 | 0.921 | 0.989 | 1.000 | 1.000 |
| 50 | 0.077 | 0.063 | 0.054 | 0.057 | 0.965 | 0.999 | 1.000 | 1.000 |
| 100 | 0.072 | 0.068 | 0.057 | 0.044 | 0.992 | 1.000 | 1.000 | 1.000 |
| 200 | 0.073 | 0.065 | 0.052 | 0.043 | 0.998 | 1.000 | 1.000 | 1.000 |

Table 3: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (33)- (35), i.e. the case of no common factor structure in the regressors.

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.058 | 0.050 | 0.044 | 0.045 | 0.980 | 0.999 | 1.000 | 1.000 |
| 50 | 0.059 | 0.049 | 0.046 | 0.044 | 0.998 | 1.000 | 1.000 | 1.000 |
| 100 | 0.062 | 0.046 | 0.048 | 0.040 | 0.999 | 1.000 | 1.000 | 1.000 |
| 200 | 0.071 | 0.050 | 0.046 | 0.046 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.061 | 0.052 | 0.047 | 0.047 | 0.904 | 0.976 | 1.000 | 1.000 |
| 50 | 0.062 | 0.051 | 0.049 | 0.047 | 0.978 | 1.000 | 1.000 | 1.000 |
| 100 | 0.064 | 0.048 | 0.049 | 0.041 | 0.986 | 1.000 | 1.000 | 1.000 |
| 200 | 0.073 | 0.051 | 0.046 | 0.047 | 0.998 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.070 | 0.064 | 0.056 | 0.055 | 0.611 | 0.739 | 0.991 | 1.000 |
| 50 | 0.070 | 0.057 | 0.055 | 0.052 | 0.778 | 0.966 | 1.000 | 1.000 |
| 100 | 0.067 | 0.050 | 0.053 | 0.044 | 0.791 | 0.980 | 1.000 | 1.000 |
| 200 | 0.074 | 0.054 | 0.047 | 0.048 | 0.938 | 0.999 | 1.000 | 1.000 |

Table 4: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (36), i.e. the case of a pure factor model for $y_{it}$.

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.044 | 0.037 | 0.037 | 0.030 | 0.915 | 0.959 | 0.988 | 0.996 |
| 50 | 0.038 | 0.034 | 0.036 | 0.033 | 0.993 | 0.999 | 1.000 | 1.000 |
| 100 | 0.042 | 0.041 | 0.036 | 0.032 | 1.000 | 1.000 | 1.000 | 1.000 |
| 200 | 0.046 | 0.043 | 0.038 | 0.036 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.047 | 0.036 | 0.037 | 0.030 | 0.773 | 0.860 | 0.935 | 0.970 |
| 50 | 0.040 | 0.035 | 0.036 | 0.033 | 0.957 | 0.987 | 0.998 | 1.000 |
| 100 | 0.042 | 0.042 | 0.037 | 0.033 | 0.999 | 1.000 | 1.000 | 1.000 |
| 200 | 0.047 | 0.044 | 0.038 | 0.037 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.054 | 0.042 | 0.038 | 0.032 | 0.467 | 0.525 | 0.635 | 0.733 |
| 50 | 0.047 | 0.038 | 0.039 | 0.035 | 0.703 | 0.822 | 0.912 | 0.962 |
| 100 | 0.049 | 0.047 | 0.038 | 0.035 | 0.967 | 0.994 | 0.999 | 1.000 |
| 200 | 0.055 | 0.050 | 0.041 | 0.040 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 5: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (33)-(34).

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.044 | 0.039 | 0.045 | 0.039 | 0.954 | 0.989 | 0.998 | 0.987 |
| 50 | 0.044 | 0.042 | 0.038 | 0.036 | 0.998 | 1.000 | 1.000 | 0.999 |
| 100 | 0.042 | 0.038 | 0.040 | 0.038 | 1.000 | 1.000 | 1.000 | 1.000 |
| 200 | 0.043 | 0.047 | 0.041 | 0.036 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.045 | 0.041 | 0.046 | 0.041 | 0.863 | 0.933 | 0.979 | 0.995 |
| 50 | 0.045 | 0.043 | 0.039 | 0.037 | 0.978 | 0.996 | 1.000 | 1.000 |
| 100 | 0.044 | 0.039 | 0.040 | 0.038 | 0.999 | 1.000 | 1.000 | 1.000 |
| 200 | 0.048 | 0.048 | 0.043 | 0.037 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.052 | 0.049 | 0.050 | 0.043 | 0.561 | 0.646 | 0.780 | 0.859 |
| 50 | 0.047 | 0.049 | 0.042 | 0.039 | 0.809 | 0.894 | 0.960 | 0.991 |
| 100 | 0.052 | 0.042 | 0.043 | 0.040 | 0.978 | 0.997 | 1.000 | 1.000 |
| 200 | 0.058 | 0.052 | 0.044 | 0.039 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 6: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (33)-(35), i.e. the case of no common factor structure in the regressors $x_{it}$.

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.048 | 0.054 | 0.053 | 0.056 | 0.973 | 0.994 | 0.998 | 1.000 |
| 50 | 0.042 | 0.041 | 0.046 | 0.049 | 0.999 | 1.000 | 1.000 | 1.000 |
| 100 | 0.039 | 0.043 | 0.044 | 0.041 | 1.000 | 1.000 | 1.000 | 1.000 |
| 200 | 0.043 | 0.040 | 0.039 | 0.040 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.049 | 0.056 | 0.054 | 0.057 | 0.904 | 0.955 | 0.988 | 0.997 |
| 50 | 0.046 | 0.042 | 0.047 | 0.050 | 0.989 | 0.997 | 1.000 | 1.000 |
| 100 | 0.042 | 0.044 | 0.046 | 0.041 | 1.000 | 1.000 | 1.000 | 1.000 |
| 200 | 0.045 | 0.041 | 0.040 | 0.041 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.060 | 0.064 | 0.057 | 0.058 | 0.646 | 0.729 | 0.830 | 0.905 |
| 50 | 0.053 | 0.047 | 0.050 | 0.051 | 0.869 | 0.934 | 0.977 | 0.995 |
| 100 | 0.049 | 0.052 | 0.049 | 0.043 | 0.991 | 0.999 | 1.000 | 1.000 |
| 200 | 0.052 | 0.046 | 0.044 | 0.043 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 7: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (36), i.e. the case of a pure factor model for $y_{it}$.

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| | $\sigma_\epsilon^2 = 1/3$ | | | | $\sigma_\epsilon^2 = 1/3$ | | | |
| 30 | 0.07 | 0.058 | 0.059 | 0.054 | 0.975 | 0.998 | 1.000 | 1.000 |
| 50 | 0.072 | 0.068 | 0.053 | 0.049 | 0.991 | 0.999 | 1.000 | 1.000 |
| 100 | 0.08 | 0.064 | 0.054 | 0.050 | 0.998 | 1.000 | 1.000 | 1.000 |
| 200 | 0.079 | 0.064 | 0.054 | 0.046 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1/2$ | | | | $\sigma_\epsilon^2 = 1/2$ | | | |
| 30 | 0.075 | 0.060 | 0.058 | 0.051 | 0.901 | 0.983 | 0.999 | 1.000 |
| 50 | 0.073 | 0.069 | 0.054 | 0.049 | 0.952 | 0.997 | 1.000 | 1.000 |
| 100 | 0.077 | 0.064 | 0.053 | 0.048 | 0.983 | 1.000 | 1.000 | 1.000 |
| 200 | 0.077 | 0.063 | 0.054 | 0.044 | 0.996 | 1.000 | 1.000 | 1.000 |
| | $\sigma_\epsilon^2 = 1$ | | | | $\sigma_\epsilon^2 = 1$ | | | |
| 30 | 0.088 | 0.073 | 0.069 | 0.062 | 0.646 | 0.846 | 0.986 | 1.000 |
| 50 | 0.079 | 0.074 | 0.060 | 0.055 | 0.723 | 0.917 | 0.997 | 1.000 |
| 100 | 0.080 | 0.066 | 0.055 | 0.052 | 0.820 | 0.972 | 1.000 | 1.000 |
| 200 | 0.079 | 0.063 | 0.055 | 0.049 | 0.902 | 0.993 | 1.000 | 1.000 |

Table 8: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (33)- (34). The first-step estimator is the one proposed Song (2013).

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| 30 | 0.103 | 0.087 | 0.088 | 0.100 | 0.838 | 0.905 | 0.966 | 0.994 |
| 50 | 0.090 | 0.083 | 0.078 | 0.074 | 0.956 | 0.988 | 0.999 | 1.000 |
| 100 | 0.081 | 0.071 | 0.063 | 0.065 | 0.999 | 1.000 | 1.000 | 1.000 |
| 200 | 0.072 | 0.061 | 0.063 | 0.054 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 9: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : \gamma_i = \gamma$. The test in (3) is computed using the HAC estimator of $\Sigma_{\gamma i}$ in (7).

| | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|
| | $T$ | | | | $T$ | | | |
| $n$ | 30 | 50 | 100 | 200 | 30 | 50 | 100 | 200 |
| 30 | 0.118 | 0.121 | 0.144 | 0.159 | 0.976 | 0.985 | 0.985 | 0.987 |
| 50 | 0.08 | 0.063 | 0.069 | 0.082 | 0.985 | 0.991 | 0.992 | 0.995 |
| 100 | 0.05 | 0.046 | 0.036 | 0.04 | 0.997 | 0.998 | 0.999 | 0.999 |
| 200 | 0.061 | 0.039 | 0.036 | 0.036 | 0.999 | 1.000 | 1.000 | 1.000 |

Table 10: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$. The test in (4) is computed using the HAC estimator of $\Sigma_{ft}$ in (10).